**Finance and Economics Discussion Series**
**Divisions of Research & Statistics and Monetary Affairs**
**Federal Reserve Board, Washington, D.C.**

# The Local Impact of Containerization

**Leah Brooks, Nicolas Gendron-Carrier, and Gisela Rua**

**2018-045**

Please cite this paper as:
Brooks, Leah, Nicolas Gendron-Carrier, and Gisela Rua (2018). "The Local Impact of Containerization," Finance and Economics Discussion Series 2018-045. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2018.045.

Working Paper

**The Local Impact of Containerization**

Leah Brooks
Trachtenberg School of Public Policy and Public Administration
George Washington University

Nicolas Gendron-Carrier
Department of Economics
University of Toronto

Gisela Rua
Division of Research and Statistics
Board of Governors of the Federal Reserve System

Friday 1st June, 2018

Thanks to Elliot Anenberg, Nathaniel Baum-Snow, Paul Carrillo, Jonathan Dingel, Jessie Handbury, Thomas Holmes, Peter Morrow, Luu Nguyen, Justin Pierce, Lukas Püttmann, Jordan Rappaport, and Matthew Turner for many helpful comments and discussions. We are also very appreciative to research assistants James Calello, Adrian Hamins-Puertolas, Arthi Rabbane, Alex Severn, and Daniel Walker who helped with data entry. We are grateful to Matthew Turner and Gilles Duranton for sharing the 1956 County Business Patterns data. Finally, we thank Nora Brooks, retired reference librarian and Leah's mother, whose combination of digital and pre-digital skills uncovered the 1948 port-level trade data.

# The Local Impact of Containerization

We investigate how containerization impacts local economic activity. Containerization is premised on a simple insight: packaging goods for waterborne trade into a standardized container makes them dramatically cheaper to move. We use a novel cost-shifter instrument – port depth pre-containerization – to contend with the non-random adoption of containerization by ports. Container ships sit much deeper in the water than their predecessors, making initially deep ports cheaper to containerize. Consistent with New Economic Geography models, we find that counties near container ports grow an additional 70 percent from 1950 to 2010. Gains predominate in counties with initially low population density and manufacturing.

Leah Brooks
Trachtenberg School of Public Policy
George Washington University
805 21st St. NW, Room 601F
Washington, DC 20006
lfbrooks@gwu.edu

Nicolas Gendron-Carrier
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON, Canada, M5S 3G7
nicolas.gendron.carrier@mail.utoronto.ca

Gisela Rua
Division of Research and Statistics
Board of Governors of the Federal Reserve System
20th and C Streets, Mail Stop 82
Washington, DC 20551
gisela.rua@frb.gov

Underlying the second wave of globalization following World War II is a vast improvement in the ability to transport goods. New York City's Herald Square Macy's now finds it cheaper to source a dress from Malaysia than from the city's own rapidly disappearing garment district (Levinson, 2008, p. 3). This decline in the importance of physical distance owes much to the development and rise of containerization. Containerization, which took off in the early 1960s, is premised on a simple insight: packaging goods for waterborne trade into a standardized container makes them cheaper to move. Containerization simplifies and speeds packing, transit, pricing, and the transfer from ship to train to truck. It also limits previously routine and lucrative pilferage.

These cost declines have yielded sea changes in trade. From the advent of containerization in 1956 to 1981, containerization caused international trade to grow by more than 1,000 percent (Bernhofen et al., 2016). Containerized cargo now accounts for over half of global non-commodity trade (United Nations Conference on Trade and Development, 2013).

In this paper, we use novel data and a new identification strategy to understand how the drastic decline in trade cost brought by containerization impacts local economic activity. We address the non-random adoption of the new shipping technology by ports with a novel cost-shifter instrument: port depth pre-containerization. Such a instrument isolates exogenous, cost-driven port containerization from adoption due to local demand. Because container ships sit much deeper in the water than their predecessors, they require deeper ports in which to dock. Dredging a harbor to increase depth is possible, but it is extremely costly.

We find that the cost advantage conferred by a deep harbor in the pre-containerization era makes a port more likely to containerize. To ensure that the instrument works through the cost of supplying a container port and not through a port's initial competitive advantage, we limit the instrument to only ports that are "very deep" pre-containerization. Intuitively, these ports had a depth beyond any pre-containerization advantage.

To undertake this analysis, we combine a large variety of data sources for the period 1910 to 2010. We use US counties as our unit of analysis, with county-level demographics and income from the Decennial Census (1910 to 2010) and employment and payroll data from the County Business Patterns (1956 and 1971 to 2011). We supplement these data with information on ports from 1953 and 2014, containerization adoption, and port-level foreign trade in the pre-containerization era. To measure contemporaneous alternative transportation, we use newly digitized highway and rail routes circa 1950.

We assess the economic impact of containerization's most important feature—the reduction in trade costs—through the lens of a New Economic Geography (NEG) model (Helpman, 1995; Redding and Sturm, 2008; Redding and Rossi-Hansberg, 2017). In these models, agglomeration and dispersion forces account for the spatial distribution of economic activity, and population moves in response to changes in real wages. A NEG model predicts that containerization's reduction in trade costs causes an increase population near containerized ports, and that this effect decays as distance from the port increases. Containerization has an ambiguous impact on nominal wages, depending on the balance between productivity gains from access to a larger market and greater competition from lower-priced distant firms.

Our findings are consistent with these theoretical predictions. From 1950 to 2010, our instrumental variable estimates report that counties within 100 km of a container port experience population gains of an additional 53 percentage points, which corresponds to an increase in population of about 70 percent over the 60 year period. We find smaller, but still economically meaningful, gains in employment, and no substantive change in nominal wages. These gains in population and employment dissipate with distance from the port, and are indistinguishable from zero beyond roughly 300 km.

We find that measures of initial land values mediate gains to containerization. Containerization requires large extensions of land, as port activity shifts from water-based finger piers to giant cranes and vast marshalling yards. New Economic Geography models also predict that initially smaller locations experience a proportionally larger increase in access to new markets. Consistent with both the NEG intuition and the fact that container ports are cheaper and easier to develop in initially low land value areas, we find that gains from containerization, in percentage terms, are concentrated in places where we expect low pre-containerization land values: counties with initially lower population density and manufacturing employment.

Our paper adds to several literatures. First, our findings contribute to the debate on the impact of globalization on economic activity. Following Romer and Frankel (1999), a large literature has emerged to understand how improved access to international markets affects country level outcomes such as GDP (e.g. Feyrer, 2009a,b; Pascali, 2017).[1] Our paper contributes to this literature by looking at how the reduction in trade costs brought by containerization affects the spatial distribution of economic activity within countries. In doing so, our results shed light on the potential uneven impacts of globalization. To the best of our knowledge, only one other paper isolates the causal effects of globalization on local economic activity: Campante and Yanagizawa-Drott (2017). These authors exploit constraints on the capacity of airplanes to fly long distances to obtain a source of exogenous variation in access to international markets at the city level. As do we, they find large positive effects of access to international markets on local economic activity.[2]

[1] Most papers in this literature find that improved access to international markets has large positive effects on GDP, with the exception of Pascali (2017) who documents mainly negative effects. Pascali (2017) is particularly related to our paper in that he exploits a major improvement in the shipping technology—the advent of the steamship—to examine how a decline in international transportation costs impacts economic activity.

[2] Our paper also complements a growing literature in international trade that looks at the impact of trade shocks on local labour markets (e.g. Topalova, 2010; Autor et al., 2013; Kovak, 2013). These papers

Second, our paper contributes to a growing academic literature investigating the consequences of improvements in transportation infrastructure on local economic activity (Baum-Snow, 2007; Michaels, 2008; Duranton and Turner, 2012; Donaldson and Hornbeck, 2016). These studies examine how investments in highways and railways have shaped the spatial distribution of economic activity within countries. Our paper is the first to study how large investments in maritime transportation infrastructure, specifically new container terminals, affect the economic conditions of target areas. Methodologically, our paper contributes a new instrumental variable strategy to contend with the non-random allocation of transportation infrastructure. Specifically, we introduce a cost-shifter instrument to obtain a source of quasi-random variation in the observed infrastructure. See Redding and Turner (2015) for a recent survey of the literature.

Finally, our work enhances the growing literature on containerization by expanding its focus beyond the shipping and trade industries. In this burgeoning literature, Rua (2014) investigates the global adoption of containerization, and Bernhofen et al. (2016) estimate its impact on world trade.[3] Hummels (2007), Bridgman (2014) and Coşar and Demir (2018) all analyze containerization's impact on shipping costs.

The remainder of this paper is organized as follows. The following section provides background on containerization, Section 3 outlines the theoretical motivation, and Section 4 discusses the data. We present empirical methods in Section 5, and results in Section 6. We conclude with Section 7.

---

compare locations within a country that have similar access to international markets but that, because of initial differences in industry composition, are differentially affected by changes in a trading partner's economic activity (e.g. China). In contrast, we control for initial differences in industry composition and compare locations that experience differential gains in access to international markets.

[3] The seminal book on this topic is Levinson (2008).

## 2 Containerization

Before goods went into the box, shipping was expensive and slow. Vessels spent weeks at ports while gangs of dockworkers handled cargo piece by piece. Port costs accounted for a sizeable share of the total cost of the movement of goods. The American Association of Port Authorities estimated that in-port costs, primarily labor, accounted for half the cost of moving a truckload of medicine from Chicago to Nancy, France in 1960 (Levinson, 2008, p. 9).

In response to these high costs, producers searched for alternatives. Trucker and entrepreneur Malcolm McLean is generally credited with being the first to match vision with reality when he moved 58 truck trailers on a ship from Newark to Houston in 1956 on the maiden container voyage.

Containerized trade relies on two key innovations. The first is the mechanization of container movements. Rather than workers with carts, specialized container cranes lift containers in and out of ships, around the port, and onto rail cars and trucks. This mechanization substantially decreased per unit labor costs, cut time in ports and made ever-larger ships viable. Today's Post-Panamax ship is more than 17 times larger than the first ship to carry container goods in 1956 (see ship sizes in Appendix Figure 1).

The second key innovation of containerization is the development of common standards for container size, stacking techniques, and grip mechanisms. These standards allow a container to be used across modes of transportation—ships, trucks, rail—and across countries. The U.S. standard for containers was adopted in the early 1960s, and the international standard followed in the late 1960s.

To achieve economies of scale, containerization requires physical changes to ports. In breakbulk ports, as cargo ports were known before the rise of containerization, ships pulled into finger piers and workers on- and off-loaded items by hand and cart. Ports were centrally located within cities and used a large amount of labor and a moderate amount of land for warehousing and storage. In contrast, containerized ports require substantially less labor per unit of weight and a much larger amount of land. Land is used both for the large cranes that move containers and for the marshalling of containers and trucks (Rua, 2014).

Despite containerization's small-scale start, it diffused extremely rapidly across the United States. The bulk of domestic containerization adoption occurred in the 1960s, as shown in Figure 1a, which reports the total number of US containerized ports by year. In the early 1960s, the cost decreases from containerization were perceived as primarily a domestic benefit, or following Benjamin Chinitz, "a trend far more advanced in domestic waterhauls than in foreign trade" (Chinitz, 1960, p. 85). Containerization adoption in the United States continued at a slower pace throughout the 1970s and 1980s and plateaued thereafter.

Post-containerization, the distribution of dominant ports has shifted. Of the ten largest ports before containerization (in 1955, measured in terms of international trade), two never containerized: New York (Manhattan), NY and Newport News, VA. In fact, the Port of Manhattan, the largest in the world in 1956, no longer exists as a freight port. Of today's 25 largest ports, four did not rank in the pre-containerization top 25. Only two of the modern ten largest ports were in the pre-containerization top ten: Norfolk, VA and Los Angeles, CA.[4]

Adoption of containerization in the rest of the world followed a similar pattern, roughly one decade delayed. Figure 1b shows that the majority of containerization outside the US occurred in the 1970s (see also Rua (2014)). The pace of adoption in the US and across the world is consistent with the initial pattern of containerized trade. Until at least the mid 1960s, containerized trade was primarily domestic. The first international

[4]See Kuby and Reid (1992) on port concentration.

container service did not begin until 1966, nearly a decade after the first US shipment.

Containerized trade is now central to the global economy. Bernhofen et al. (2016) estimate that containerization caused international trade to grow by more than 1,000 percent over the 15 years following 1966. As of 2013, containerized trade accounted for over half of global non-commodity trade (United Nations Conference on Trade and Development, 2013).[5]

The literature credits containerization with substantially decreasing the cost of waterborne trade. While Bridgman (2014) and Hummels (2007) note only a small decline in shipping rates, traditional measures of shipping costs understate the true cost advantage yielded by containerization. Containerization cuts the time ships spend at port and thus the total time in transit. Hummels and Schaur (2013) estimate that each day in transit is worth between 0.6 to 2.1 percent of the value of the good, showing that the time benefits of containerized shipping are non-negligible. In addition, losses to pilferage plummeted with containerization. Wilson (1982) estimates loses to pilferage at roughly 25 percent in the breakbulk era, and near zero in the container era.[6] Finally, containers ease logistics costs by protecting goods from unintentional damage and allowing different kinds of goods, with different destinations, to be shipped together (Holmes and Singer, 2017). Using 2013 export transaction data for Turkey, Coşar and Demir (2018) find that containerization decreases variable shipping costs between 16 to 22 percent.

[5]While containers are appropriate for carrying many goods, as diverse as toys and frozen meat, some goods are not yet containerizable. Both "non-dry cargo" and "dry-bulk commodities" such as oil, fertilizers, ore, and grain cannot be shipped inside "the box."

[6]It is therefore no surprise that Scottish whiskey bound for US markets was on the first international container trip (Levinson, 2008, p. 165).

# 3 Theoretical Motivation

We now turn to the theoretical literature to frame our empirical work and understand containerization's potential impact. Containerization's most important feature is the reduction in waterborne transit costs it generates. Because almost all goods transported by water require additional land-based movement, reductions in trade costs due to containerization are largest in percentage terms at the port and decay as distance to the port increases.

We assess the impact of this reduction in trade costs through the lens of a standard New Economic Geography model (e.g. Helpman, 1995; Redding and Sturm, 2008; Redding and Rossi-Hansberg, 2017). In this class of models, agglomeration and dispersion forces explain the uneven distribution of economic activity across space, resulting in particular from people moving in response to changes in real wages. Variation in real wages typically results from changes in (1) nominal wages, (2) the cost of living, and (3) land prices.

Containerization's reduction in trade costs has three main short run effects. First, when firms produce differentiated products and consumers love variety, locations with lower trade costs become more attractive to consumers. These locations offer a greater variety of goods at lower prices, reducing the cost of living and increasing real wages.

Second, if there are increasing returns to scale in production, a reduction in trade costs also increases the profitability of firms because firms can access a larger market for their products. This "home market effect" yields an increase in nominal wages and, therefore, an increase in real wages.

Third, due to increased trade, firms encounter more lower-priced competitors. This heightened competition, known as the "market crowding effect," acts as a dispersion force and causes both nominal and real wages to decline.

If there are gains from trade, as New Economic Geography models assume, the cost of living effect and the home market effect should dominate the market crowding effect. Thus, we expect a short run increase in real wages in locations near container ports.[7]

In the long run, however, higher real wages should attract people to locations near container ports. As population increases, land prices rise, in turn lowering real wages. Migration ceases when real wages equalize across space.

Since the containerization-induced reduction in trade costs declines with distance from the port, we anticipate that the impact of containerization on population is greatest in places near container ports and declines as distance to the port increases.

The simplest New Economic Geography framework, outlined above, assumes that places are all ex-ante homogeneous. However, an extension to the basic framework can allow the same shock to impact cities unevenly, as a function of the city's initial characteristics. In the empirical section, we consider variation in both initial population and land values. Firms in initially less populous cities rely more heavily on the demand from non-local consumers. We therefore expect containerization to have a larger impact, in percentage terms, in initially smaller cities relative to initially larger cities.

We also expect containerization to have an uneven effect based on pre-containerization land prices. Because container ports require large swaths of land for giant cranes and extensive marshalling yards, rather than the water-based finger piers of the breakbulk era, container ports may be more viable in locations with initially low land value. However, as local productivity shocks are ultimately capitalized into the value of land (Moretti, 2011), low land value cities tend also to be small cities, all else equal. Thus, empirically, the distinction between being initially low population and initially low land value may not be empirically visible.

[7]Note that even if there are gains from trade, the net effect on nominal wages is ambiguous because the home-market effect and the market-crowding effect go in opposite directions.

In sum, New Economic Geography models predict that containerization's reduction in trade costs causes population to increase near container ports. This effect diminishes as distance to the container port increases. Containerization's net effect on nominal wages remains theoretically ambiguous because the productivity gains associated with access to a larger market may be offset by the intensified competition from distant firms. Finally, for a given distance to a container port we anticipate greater population growth in initially smaller cities. These smaller cities receive a proportionately larger increase in access to new markets and have relatively cheap land, which is key to container port development.

# 4 Data

To study the impact of containerization on local economic activity, we construct a county-level panel dataset that includes population and employment information, as well as proximity to port and port characteristics. This section gives an overview of the data, and we present full details in the data appendix.

Our sample frame is the Decennial Census, for the years 1910 to 2010.[8] We assemble a time invariant panel of counties by aggregating 1950 counties to their 2010 counterparts and by dropping a very few counties with large land area changes. From 1910 to 2010 we observe population; and from 1950 to 2010 income and demographic characteristics. We also observe total employment, total payroll, and employment and payroll by industry from the County Business Patterns from 1956 and then annually 1971 to 2011.[9] We omit Alaska from our analysis because its administrative districts in 1950 do not correspond

[8]For the 2010 sample, we use the Decennial Census for population figures and the American Community Survey (years 2008–2012) for other demographic covariates.

[9]We are very appreciative of digitized 1956 County Business Patterns from Matt Turner and Gilles Duranton. See the data appendix for more information about these data.

to modern counties. This yields 3,023 counties with complete data.[10]

To this sample frame, we add port attribute data. Our universe of ports is all ports that existed in either 1953 or 2015, as defined by the 1953 and 2015 *World Port Index*. For each port, we observe its location (latitude and longitude), size (in four discrete categories), and depth (in eight discrete categories). We gather the year of first containerization from the *Containerisation International Yearbook*, volumes 1968 and 1970 to 2010.[11] We also observe 1948 and 1955 international trade in dollars by port from the Census Bureau's Foreign Trade Statistics. We associate each county with a vector of ports and port characteristics, which include the distance from each county to each port, the number of nearby 1953 ports, the maximal depth of nearby ports in 1953, and the total value of international trade at nearby ports in 1948 and 1955.[12]

We also include variables that characterize the state of the transportation network now and at the advent of containerization (c. 1957 for highway and c. 1960 for rail). We measure total rail kilometers, highway kilometers, and waterway kilometers in each county, per square kilometer of each county's area.

In addition to these detailed US data, we construct a less detailed panel dataset of world cities. The sample frame for world cities is the United Nation's 2014 Revision of World Urbanization Prospects. This dataset contains all 1,692 urban agglomerations with populations exceeding 300,000 at any time between 1950 and 2014. By construction, this sample over-represents fast growing cities that were small in 1950 but grew rapidly in the second half of the twentieth century. To mitigate this sampling issue, we restrict the sample to cities with population over 50,000 in 1950, yielding a world panel of 1,051

[10]Estimations using County Business Patterns data use a slightly smaller sample because the provider suppresses data for counties under certain conditions; see data appendix for complete details.

[11]For the purposes of this paper, and consistent with the industry definition, we call a port "containerized" when it has special infrastructure and equipment to handle containers. Specifically, the port has invested in equipment to handle shipping containers which enables their movement in and out of ship and onto a train or a truck.

[12]We calculate all distances from the county centroid.

cities.

# 5 Empirical Methods

We now turn to our empirical strategy for estimating the causal effect of containerization on local economic activity. We first present a difference-in-differences framework to analyze the impact of proximity to a containerized port on economic activity and illustrate its strengths. We then discuss remaining concerns with causality, followed by a motivation for and details about our instrumental variable strategy.

## 5.1 Difference-in-Differences

Our goal is to understand how local economic activity responds to the advent of containerization. Specifically, we test the theoretical predictions that population and employment increase in locations close to container ports and that these gains attenuate with distance from the port. We also test whether percentage gains are larger in locations with initially low land values, all else equal. Empirically, we ask whether county proximity to a containerized port is associated with changes in key economic outcomes, conditional on a host of covariates. We estimate

$$\Delta \ln(y_{i,t}) = \beta_0 + \beta_1 \Delta C_{i,t} + \beta_2 X_i + \Delta\epsilon_{i,t} \ , \tag{1}$$

where $i \in I$ indexes counties and $t \in T$ indexes years. Our primary dependent variable, $y_{i,t}$, is population. We also investigate the impact that containerization has on nominal wages, industrial composition, and income. The operator $\Delta$ denotes long run differences, so that $\Delta \ln(y_{i,t}) = \ln(y_{i,t}) - \ln(y_{i,1950})$.[13] Capital letters denote vectors.

[13]When we use County Business Patterns data, the initial year is 1956.

Our key explanatory variable is an indicator for proximity to a containerized port at time $t$, $\Delta C_{i,t}$, which is equivalent to $C_{i,t}$, as no containerized ports existed in 1950 ($C_{i,1950} = 0 \ \forall i \in I$). We allow for potential non-linear impacts of proximity to a containerized port by using indicator variables for port proximity by distance bin. Figure 3a shows this parameterization. Counties in the darkest blue are located within 100 km of a containerized port, counties in mid-blue are between 100 and 200 km from a containerized port, counties in light blue are between 200 and 300 km from a containerized port, and counties in light pink are more than 300 km away from a containerized port.

Mathematically, we parameterize proximity to a containerized port as

$$\beta_1 \Delta C_{i,t} \equiv \sum_{d \in D} \beta_{1,d} \mathbb{1}\{\text{Closest containerized port is between } d_1 \text{ and } d_2 \text{ km}\}_{i,t}, \qquad (2)$$

where $d \in D$ are a set of distance bins of $\{0-100, 100-200, 200-300\}$ kilometers. We interpret $\beta_{1,\{0-100\}}$ as the percentage change in the dependent variable for counties within 100 km of a containerized port relative to counties more than 300 km away from a containerized port, conditional on covariates. Coefficients $\beta_{1,\{100-200\}}$ and $\beta_{1,\{200-300\}}$ refer to the remaining distance bins.

Theory suggests that population increases in counties proximate to container ports ($\beta_1 > 0$). In addition, standard New Economic Geography models predict that containerization's impact attenuates with distance from the port, so that $\beta_{1,\{0-100\}} > \beta_{1,\{100-200\}} > \beta_{1,\{200-300\}}$.[14] However, theory does not clearly predict where the impact of containerization stops, so this bound of 300 km comes from the data (see a more detailed discussion on this in Section 6.2, footnote 24).

To establish the causal effects of containerization on local economic activity, we must contend with the non-random assignment of containerized ports to counties. The

[14]This framework does not allow us to distinguish between growth and reallocation. See footnote 21 for a discussion of the magnitude of reallocation required for growth to be negligible.

difference-in-differences specification in Equation (1) goes some way to this end by netting out any time-invariant county-specific characteristics correlated with the location of containerized ports. Such characteristics include geography, proximity to population centers, climate, and historical antecedents for the location of particular industries. This method also nets out any national changes that impact all counties equally from 1950 to 2010.

In the event that county proximity to a containerized port is also a function of time-varying county attributes, we also include a vector of baseline covariates, $X_i$. Including initial covariates in the difference-in-differences model is akin to allowing for differential trends in the dependent variable by the initial covariates. We list these in greater detail in Section 6, but $X_i$ includes regional fixed effects, distance to the ocean, measures of geographic proximity to ports in 1953, the extent of the initial transportation network, initial demographic characteristics, initial industry mix, and pre-1950 county population. We cluster standard errors throughout at the 2010 commuting zone to account for spatial dependence in the error. A commuting zone is a grouping of counties that approximate a local labor market. The average commuting zone includes 4.4 counties.[15]

This empirical strategy yields a causal estimate of the effect of proximity to a containerized port on local economic activity when proximity to a containerized port is uncorrelated with the error term. This is equivalent to saying that $\beta_1$ can be interpreted as a causal estimate when proximity to a containerized port is randomly assigned, conditional on time-invariant county-level factors and the included initial covariates. Because we include a host of initial period covariates, these estimates cannot be driven by, for ex-

[15] We have also made standard error estimates with the spatial HAC method, using radii of 100, 200 and 300 km. Because these standard errors are in general smaller than those using commuting zones, and because these spatial standard errors are not (to the best of our knowledge) yet available for the instrumental variable case, we use commuting zone clustering throughout. Even in principle, commuting zone clustering may be preferred, as commuting zone counties are linked by economic activity and therefore likely to be spatially correlated. In contrast, counties within a fixed radius may be less likely to be related in an economically meaningful way.

ample, regional trends in population growth, or differential population growth related to proximity to the coast.

To test the predictions that gains vary by initial conditions, we introduce an interaction term that allows $\beta_1$ to vary below the median of a given covariate. Call this covariate $h_i$ and let $H_i = 1$ when $h_i < \text{median}(h_i)$ and 0 otherwise.[16] We therefore modify Equation (1):

$$\Delta \ln(y_{i,t}) = \gamma_0 + \gamma_1 \Delta C_{i,t} + +\gamma_2 \Delta C_{i,t} * H_i + \gamma_3 X_i + \gamma_4 H_i + \Delta\epsilon_{i,t} \ . \tag{3}$$

Now $\gamma_1$ reports the average impact of proximity to a container port on population growth, and $\gamma_2$ reports whether there is any incremental population gain or loss in counties when $h_i$ is below the median. We expect containerization induced population growth to be larger, in percentage terms, in locations with low initial population and low initial land values. We therefore anticipate $\gamma_2 > 0$ when $h_i$ is a measure of initial land values or population.

While both equations (1) and (3) net out county-specific time-invariant factors as well as trends by initial conditions – including distance to the ocean and initial share of employment in manufacturing – it may still be the case that an element in the error $\Delta\epsilon_{i,t}$ remains correlated with both containerization and the outcome variable of interest. For example, if counties near container ports were more likely to specialize in an agricultural commodity that became tradeable since the 1950s, we could conflate local economic growth due to the increase in the trade of the agricultural commodity with local economic growth related to containerization.

[16] $H_i$ relative to the overall distribution and $H_i$ relative to the treated distribution are both of interest. We consider both empirically; in practice the difference in estimates is quite small.

## 5.2 Instrumental Variables

To address this type of concern – and any other remaining non-randomness in the assignment of containerized ports to counties – we use proximity to a very deep port in 1953, $Z_i$, as an instrument for proximity to a containerized port, $\Delta C_{i,t}$. Specifically, we instrument proximity to a containerized port with proximity to initially very deep ports as

$$\Delta C_{i,t} = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_i + \Delta \eta_{i,t} \ , \tag{4}$$

where $\alpha_1 Z_i$ is

$$\alpha_1 Z_i \equiv \sum_{d \in D} \alpha_{1,d} \mathbb{1}\{\text{Closest very deep port in 1953 is between } d_1 \text{ and } d_2 \text{ km}\}_i. \tag{5}$$

Thus, we have three potentially endogenous variables and three instruments. For the interaction specification in Equation (3), we use both proximity to a very deep port, $Z_i$, and that proximity interacted with being below the median of a given covariate, $Z_i * H_i$, as instruments—so, six instruments overall.

There are two requirements for the instrument to yield a causal estimate of proximity to a containerized port on local economic activity. The first is a strong relationship between proximity to a containerized port and proximity to a very deep port in 1953. The second requirement is that, conditional on covariates, proximity to a very deep port in 1953 is uncorrelated with unobserved determinants of changes in local economic activity from 1950 to period $t$. In other words, proximity to a very deep port in 1953 impacts changes in local economic activity only through its impact on proximity to a containerized port, conditional on covariates ($\text{Cov}(Z_i, \Delta \epsilon_{i,t}) = 0$). We discuss each of these requirements in turn.

First, we anticipate that proximity to a containerized port should be strongly related to proximity to a very deep port in 1953 because container ships require deeper ports than their predecessors. As Appendix Figure 1 illustrates, container ships are much larger than their predecessors and larger ships sit deeper in the water and thus require greater depth to navigate and dock.

It is possible, but quite expensive, to drill, blast or dredge an initially shallow port sufficiently deep to accept container ships. Given enough money and sufficiently lax environmental regulation, a harbor can arguably be made arbitrarily deep. However, port depth is only malleable at great cost. Therefore, initially deep ports have a competitive advantage when technology changes to favor very deep ports. This inability of ports to adjust equally is confirmed by Broeze, who notes that while "ship designers [keep] turning out larger and larger vessels," and "the engineering limits of port construction and channel deepening have by no means been reached[, t]his, however, may not be said of the capacity of all port authorities to carry the cost of such ventures" Broeze (2002, pp. 175–177). Thus, initial port depth is a key component of the cost of converting a breakbulk port into a containerized port.

Our instrument is therefore analogous to a cost shifter instrument often used in the industrial organization literature. Port depth should affect the supply of ports after the advent of containerization, but have no effect on the demand for ports.

The intuition that port depth is a key driver of containerization is borne out in practice by containerization's pattern of adoption. Figure 2a shows the likelihood that a county becomes proximate to (within 300 km of) a containerized port over time by the maximal depth of ports within 300 km of the county in 1953.[17] Thick lines indicate depths we consider "very deep."

It is immediately clear that proximity to deep ports in 1953 is a strong predictor of

[17]We use depth of the wharf in 1953 as our measure of pre-containerization port depth. Results are robust to using anchorage and channel depth, which the *World Port Index* also reports.

proximity to a containerized port at time $t$. Counties within 300 km of a port with depth greater than 40 feet are always within 300 km of a containerized port by the end of the sample period, as are almost all counties with 300 km of a port 35 to 40 feet deep. Roughly 20 percent of counties within 300 km of a port with depth between 25 and 35 feet are not near a containerized port by the end of the sample period. For counties within 300 km of less deep ports, however, containerization is decidedly not a certainty. Indeed, counties near initially shallow ports—those less than 20 feet deep—are never within 300 km of a containerized port.

An alternative way to view the strength of our instrument is to compare Figures 3a and 3b. The top panel is the map of US counties, where treated counties are blue and deeper blue indicates greater proximity to a containerized port. The bottom panel repeats this map, but re-colors treated counties in green when the instrument predicts treatment. "Predicting treatment" means that a county is both between $d_1$ and $d_2$ km from the nearest containerized port in 2010 and between $d_1$ and $d_2$ km from the nearest very deep port in 1953. This picture demonstrates that while the instrument frequently fails to predict treatment in the Midwest, it predicts treatment quite accurately on the ocean coasts.[18]

Given this evidence of a strong relationship between the endogenous variables and the instruments, we now turn to the second condition for instrument validity—that proximity to a very deep port in 1953 affects local economic activity only through its impact on proximity to a containerized port.

A key concern with the instrument is that proximity to a deep port may explain changes in county economic activity even before containerization. This is surely true, as ports have long been engines of growth. For this reason, rather than rely on the full distribution of port depth, we use an indicator variable for a county being proximate to

[18]We address this case where the instrument fails to predict treatment in Section 6.2.

a very deep port pre-containerization. Specifically, we call a port "very deep" when it is 30 feet or more deep in 1953. We choose this depth cut-off because the historical record indicates no perceived advantage to depth greater than 30 feet in the pre-containerization era.

Before containerization, while port depth conveyed some advantage, it was not particularly useful for a port to be very deep given the draft of breakbulk ships. This is clear even from how data on port depth was collected. The 1953 *World Port Index*'s deepest category is "40 feet and above," while the deepest category in the 2015 *World Port Index* is "76 feet and over." Thus, intuitively, our instrument measures how much more likely a county is to become proximate to a container port if it is proximate to a very deep port in 1953, conditional on initial covariates. Our specification includes covariates that allow for differential growth trends in the dependent variable by the number of ports in 1953 within 300 km in 100 km bins and the values of international trade at these ports in 1955, also measured in 100 km bins. Therefore, the instrument captures the impact of proximity to an initially very deep port above and beyond proximity to many ports in 1953 and to high value ports in 1955.

Our claim that depths beyond 30 feet were not particularly advantageous to port success is supported by a number of contemporary commentators. A 1938 monograph notes the critical 30-foot cut-off, arguing that "For the ports with which we are dealing, the 30-foot channel at low-water will be taken as the minimum standard in relation to the needs of modern ships" (Sargent, 1938).[19] However, he notes that the cost of making a channel deeper is no small endeavor: "It is a question how far the rest of the world, Europe in particular, is prepared, except in special circumstances, to face the very heavy cost of providing for the needs of the ocean mammoth" (Sargent, 1938, p. 21).

This author's focus on the irrelevance of extreme depth is not unique. Even as late as

[19]He goes on to write that in the U.S., a 35-foot draught is becoming standard (p. 21).

1952, F. W. Morgan argues in *Ports and Harbours* that beyond a certain level, depth is not a particularly useful feature of a port:

> The importance for a few ports of maintaining a ruling depth sufficient to admit the largest liners [a draft of 40 feet] emphasizes unduly their importance to the port world. A super-liner which comes into a port every few weeks will, it is true, amplify that port's tonnage figures by half a million tons or so annually. . . . The greater part of world trade by sea and the greater part of the traffic of many ports is concerned with ships of more modest size.
>
> It would certainly be possible to devise a classification of ports by the draught of ship which can be berthed in them. Halifax and Wellington would appear in the first class, and their ability to berth the largest ships is a great asset in wartime. It tells, however, only a little about their normal significance as ports. (p. 15, Morgan (1952))

Thus, pre-containerization, being very deep was not a particularly valuable port attribute.

This instrumental variables strategy implies multiple tests for validity. First, if our claims about the role of "very deep" ports are true, we should see no impact of proximity to very deep ports on population growth in the pre-containerization era. In addition, in any sub-sample where our instrument does not predict treatment, the instrument should have no direct impact on population growth. Finally, the instruments should not be correlated with potential confounders that might be in the error term. We turn to these tests in the instrumental variables results section.

# 6 Results

With this empirical framework in hand, we now turn to estimation. The first subsection reports summary statistics and the difference-in-differences results. The second subsection presents tests of instrument validity, discusses our main instrumental variable results, and assesses whether the results are robust to alternative specifications. The third subsection tests whether containerization's impact is larger, in percentage terms, in places with low initial land values.

## 6.1 Difference-in-Differences

We begin with the difference-in-differences specification to test the theoretical prediction that containerization increases local economic activity. The summary statistics in Table 1 illustrate the comparison at hand and preview the main results. The three leftmost columns report county means by distance to the nearest containerized port by 100 km bins; the fourth column shows means for all observations within 300 km of a containerized port, and the final column reports means for all other counties, which we call "never containerized." A county may appear in only one distance bin. The number of observations in the "ever" and "never" columns sum to the total sample size (final row). On average, counties near container ports have experienced about forty years of containerization.

The figures on log population in the first rows of this table clearly show that counties near containerized ports were larger pre-containerization and that counties closest to containerized ports were largest. From 1910 to 1950—the pre-containerization years—log population in counties near future containerized ports is larger and increases at a faster rate than in counties farther from future containerized ports. These differences between counties generate a possible bias in the OLS estimation that we address in the IV section.

The summary statistics also show some additional differences between counties by proximity to a containerized port. Across census regions, counties near containerized ports are over represented in the Northeast, under represented in the Midwest and West, and about proportionately represented in the South. Counties near containerized ports had a substantially larger share of workers in manufacturing in 1956, on average.

In addition, these summary statistics illustrate our main finding that counties near containerized ports grow at a faster pace after the advent containerization than the average untreated county. This relative increase is visible not only in the population data, but also in the employment and payroll per employee data from the County Business Patterns.

Moving to a regression framework, Table 2 presents difference-in-differences results, testing the prediction that proximity to a containerized port is associated with greater population growth after the advent of containerization. Column 1 presents estimates including only regional fixed effects and shows a 68 percentage point increase in population growth for counties within 100 km of a containerized port relative to counties more than 300 km away from a containerized port. This coefficient declines to 35 percentage points for counties between 100 and 200 km from a containerized port and to 24 percentage points for counties between 200 and 300 km from a containerized port.[20]

The remaining columns in this table add additional covariates. To address the concern that counties of different size may grow at different rates—especially since counties near containerized ports are uniformly initially larger—Column 2 controls for log of population in years 1920, 1930 and 1940. We also add controls for the share of population with a college degree and share African American by county, both measured as of 1950.

To isolate the impact of containerization from proximity to the coast, initial port intensity and pre-containerization port prominence, Column 3 adds additional controls. These are distance to the ocean, three variables for the number ports in 1953 within 300 km, measured in bins of 100 km, and three variables for the total value of 1955 international trade at ports within 300 km, again measured in bins of 100 km. Results

[20] In this and all estimates in this paper, we cluster standard errors by the 2010 commuting zone to account for spatial dependence across counties. See footnote 15 for more details.

decline by about one-third to one quarter, so that the gradient by distance bin is now 46, 26, and 16 percentage points, respectively.

Finally, we address the higher rates of 1956 manufacturing activity near future containerized ports, as seen in Table 1). The fourth column includes this variable and measures of the extent of pre-existing transportation networks as controls. Measures of the 1950s-era transportation are the length of highways, navigable waterways, and railways per square kilometer. These controls have little additional impact on the size of the coefficients. We now estimate 45, 25 and 13 percentage point increases in population with distance to the closest containerized port.

These results are consistent with the theoretical predictions of a standard New Economic Geography model: population increases near containerized ports and gains dissipate with distance.[21] Population increases are large and decline monotonically, but not linearly, with distance from the containerized port. We defer a detailed discussion of the magnitude of the estimates and the choice of the 300 km border until the presentation of the instrumental variables results.

## 6.2 Instrumental Variables

Although the difference-in-differences specification addresses many confounding factors potentially correlated with both proximity to a containerized port and population growth—such as past population and initial industrial mix—it is possible that some part of the error term remains correlated with the treatment. We now turn to our instrumental variables estimates. We start with the graphical reduced form intuition, proceed to instrument strength and validity, follow with instrumental variable results, and conclude

[21] Our estimation does not discriminate between growth and reallocation. In the period between 1950 and 2010, the US population roughly doubled, from about 150 to roughly 300 million. Thus, our results seem very unlikely to be driven exclusively by reallocation, as they would require approximately half of the 1950 population to relocate due to containerization.

with measures of robustness.

**Reduced Form: Relating Proximity to Very Deep Ports and Population Growth**

To give intuition for the instrument variable analysis, Figure 2b presents a graphical illustration of the reduced form regression (a regression of change in the log of population on the instrument). This figure presents the average log of population over time by initial depth category. Thick lines indicate counties within 300 km of ports that we classify as very deep in 1953; thin lines are counties within 300 km of ports less than 30 feet deep in 1953. We also include a line for counties not within 300 km of a container port. In essence, the estimation asks whether the thicker lines trend upward more after 1956 (the vertical red line) than do the thin lines. This picture shows that the thick lines of counties near very deep ports do, and that the gains are driven primarily by initially smaller counties—the beige and purple lines.

**Instrument is Strong and Unrelated to Pre-containerization Population Growth**

We already saw from Figure 2a (discussed in Section 5) that the instrument is strong. Appendix Table 1 validates this intuition, reporting coefficients for the three equations that estimate the full first-stage (one equation per distance bin). The table shows the pattern we expect if the instrument is working as we hypothesize: counties that are between $d_1$ to $d_2$ km from the closest very deep port in 1953 are more likely to be between $d_1$ to $d_2$ km from the closest containerized port in 2010. These coefficients on the diagonal are large—in the 0.5 to 0.6 range—and strongly significant. Thus, even conditional on the many covariates we use, proximity to a very deep port in 1953 remains an important predictor of proximity to a containerized port in 2010. The lowest F statistic on the instruments in any of these three equations is 22; the highest is 59. Our two-stage least squares estimates tables always report the Kleinberg-Paap F statistic, which summarizes the overall strength of the first-stage, as suggested by Sanderson and Windmeijer (2016). In our main instrumental variable estimates, this F statistic is never smaller than 21.[22]

Given that the instrument is strong, we now turn to three tests for validity. First, we examine whether proximity to a very deep port is related to pre-containerization population changes; given what we have argued, it should not be. Figure 4 shows the distribution of population change 1910 to 1950, conditional on regional fixed effects and distance to the coast. The red line shows the distribution for counties near (within 300 km) of very deep ports, and the blue line the distribution for counties far from very deep ports. These distributions are virtually indistinguishable. The 95 percent confidence interval on a dummy from a regression distinguishing between these two types of counties is small relative to the first-stage coefficients and covers zero: [-0.11,0.04]. Thus, we find little evidence that proximity to a very deep port impacts pre-containerization population growth, adding confidence in the validity of the instrument.

An additional implication of the IV framework is that, in cases where the instrument fails to predict treatment, the instrument should also be uncorrelated with the dependent variable – since the assumption underlying the instrumental variable specification is that the instrument impacts the dependent variable only through the endogenous variable. In our data, proximity to port depth fails to predict proximity to containerization in the Great Lakes region. Ports in this area were not very deep in 1953, yet regional ports did adopt containerization. If the proximity to deep ports impacts population and other outcomes only through proximity to containerization, then in cases where port depth is unrelated to containerization, it should also be unrelated to population changes (see Angrist et al. (2010), page 798).

Limiting our analysis to the roughly seven hundred counties within 300 km of the

[22]These first-stage results are also qualitatively robust to defining "very deep" as one category above (greater than 35 feet deep) or one category below (greater than 25 feet deep). The $F$ statistics are larger, and the estimates more precise, when we use the lower depth cut-off.

Great Lakes, we find a very weak relationship between proximity to port depth and proximity to containerization. Further, we see no relationship between proximity to deep ports and population growth. The coefficients on the instrument in the reduced form specification are an order of magnitude smaller than the main estimates (coefficients by distance bin are -0.040, 0.078, and 0.050) and are never different from zero. See Appendix Table 2 for complete results.

Our third test of instrument validity evaluates whether the instruments are correlated with county-level characteristics that might plausibly be in the error term. While we cannot do this for all potential confounders, we can observe whether the identifying variation—the residual from a regression of an instrumental variable on the full set of covariates from Table 2—is correlated with specific pre-treatment covariates, also conditional on covariates.

Recall that our regression specification controls for log of population in 1920, 1930, and 1940. Were the identifying variation in the instrument to be related to the log of 1910 population (conditional on covariates), this would suggest that the pre-treatment controls were not adequately capturing the historical pattern of population growth. We do not find this to be the case. We do a similar analysis for international trade at ports. Recall that the regression controls for the 1955 value of international trade flows in each of the three distance-to-containerized-port bins. If this covariate did not sufficiently control for the impact of pre-containerization port strength on population growth, we would expect that the identifying variation would be related to the 1948 value of international trade flows by distance-to-containerized-port bins, conditional on covariates.[23]

Appendix Figure 2 displays the full matrix of scatterplots showing the correlation between 1910 population and 1948 trade and the identifying variation. There are no

[23]An alternative method is to include these controls directly in the regression, and results are robust to doing so. We believe that this test, however, highlights the econometric implication of this lack of importance: that the identifying variation is not correlated with likely confounders.

significant relationships, and the largest $t$ value for any of these relationships is $2e-8$.

### Instrumental Variable Results Consistent with Difference-in-Differences Findings

Given these tests of validity, we report instrumental variable results in the right half of Table 2. The columns repeat the pattern of covariates from the OLS half of the table. The coefficients are generally quite similar, though slightly larger than the OLS in the complete specification (columns 4 and 8). Why might IV results be larger? As discussed in section 3, we expect containerization to have a larger impact on population growth in initially smaller counties. When we use the instrument to correct for endogeneity in the proximity to a containerized port, we are in principle giving more weight to initially smaller counties where the depth is the main driver of the containerization decision. As a result, coefficients in the IV regression increase.

The most complete model in column 8 shows a 53 percentage point increase in population growth over the 60 years from 1950 to 2010 for counties within 100 km of a containerized port relative to counties more than 300 km away from a containerized port. Consistent with the expected relationship between the gains to containerization and distance from the port, this coefficient declines to 29 and 20 percentage points for counties slightly farther from containerized ports.[24]

To interpret the magnitude of these results, we turn to Duranton and Turner (2012), who find that a 100% increase of a city's initial stock of highways yields a 13 percent increase in population over a 20 year period. This corresponds to an annualized increase

[24] Both here and in the OLS estimates, we compare counties within 300 km of a containerized port to all other counties. As theory does not provide guidance on the physical distance over which containerization might have a measurable impact, we turn to the data as a guide. Appendix Figure 3 shows regression coefficients from a version of Equation (1) where distance to containerized port is measured in 50 km bins. Gray bands are confidence intervals. These results show that the association between proximity to a containerized port and population growth is indistinguishable from zero at 300 km. In our main specification, we use bins of 100 km, rather than the smaller 50 km ones, to increase the power in the estimates. This is particularly important when we examine whether containerization's impacts differ by initial conditions in subsection 6.3.

of about 0.6 percent. Our findings are similar. Being within 100 km of a containerized port causes a 70 percent increase in population over a 60 year period ($\exp(.53) - 1 = .70$), implying a comparable annual growth rate. Our containerization effect is thus roughly equivalent to a doubling in the initial stock of highways in a county.[25]

**Containerization's Impact Increases Over Time**

To test for changes in the impact of containerization over time, we re-estimate Equation (1) using different final years, starting in 1970. We report coefficients from these estimations in Appendix Figure 4, which displays results decade-by-decade. Full circles are significant coefficients and hollow circles are insignificant coefficients (at the five percent level). The red line at the top reports the coefficients for counties within 100 km of a containerized port; the orange line 100 to 200, and the yellow line 200 to 300 km. Apart from a blip in 1980, counties near containerized ports have large population gains that increase over time. For example, in 1970, only 15 years after the advent of containerization, counties closest to containerized ports had grown by almost 40 additional percentage points relative to counties more than 300 km away from a containerized port. By 2010, this figure was 55. While estimates for counties farther from ports are smaller, they also follow this general pattern of increase. This increasing impact decade-by-decade may reflect the increasing size of the containerized port network, as shown in Figure 1.

**Results Robust to Additional Considerations**

We now turn to threats to identification. Rappaport and Sachs (2003) argue that coastal locations have long been associated with greater economic growth, crediting both in-

[25]Containerization required substantial investments. In the years of peak outlays from 1968 to 1973, the U.S. spent about \$2015 8 billion of public and private funds on the required port infrastructure (Kendall, 1986). This is about \$2015 1.6 billion per year, one fourth of the annualized cost of the Interstate Highway System from 1956 through 1991 (`https://www.fhwa.dot.gov/interstate/faq.cfm`, assessed on 08/21/2017).

creased productivity and, more recently, better amenities. We can interpret containerization as a productivity-enhancing mechanism that generates part of the Rappaport and Sachs result. However, our estimates show that containerization is more than just coastal proximity: our main results are little changed by the inclusion of a Rappaport and Sachs coastal indicator (Table 3 column 2).[26]

To further isolate the impact of containerization from proximity to the coast, Table 3's column 3 restricts the sample to counties within 400 km of a port in 1953. The sample size drops from 3,023 to 1,767 observations, but the coefficients decline only slightly (compare estimates to column 1, which repeats the most complete specification from Table 2). This suggests that population growth in counties near a containerized port is not driven by a comparison with slower-growing centrally located counties.

Furthermore, we know from the summary statistics in Table 1 that counties near containerized ports experience more rapid population growth pre-containerization, and this trend may have continued after 1956 irrespective of containerization. We account for this in the main estimates by including log population in 1920, 1930, and 1940. Table 3's column 4 additionally includes squares of those measures of past population, in the event that previous population impacts population growth non-linearly. Again, the estimates are little changed.

As we discussed, our instrument does not predict containerization in the Great Lakes region, which does have container ports. In addition, this region experiences the slowest population growth over our period of analysis. To allay fears that the results are driven by this potentially anomalous treatment of the Midwest, column 5 omits the Midwest region entirely, leaving 1,975 observations. Results in this column are smaller than the original specification, but the pattern of decline with distance to the closest containerized port remains. Indeed, we should expect smaller coefficients in this estimation because

[26]Rappaport and Sachs measure coast as locations within 80 km of the Great Lakes and ocean coasts.

the control group—non-Midwest, non-containerized ports—now has a higher average population growth. Note the increase in the mean of the dependent variable from 0.373 to 0.508 (final row of the table). Still, we observe a relative population increase of 44 percentage points near containerized ports, an increase of almost three-quarters of the mean.

Research in urban economics strongly suggests that growth is associated with an area's education and demographic characteristics (Moretti, 2004). Column 5 includes additional controls for the share of people 25 or older with a high school degree, the share foreign born, the number of government workers per capita, and the share age 65 and older by county. The addition of these covariates decreases the coefficients slightly, with greater impact for the category closest to containerized ports. The coefficients remain sizeable, and retain the pattern of decline with distance to containerized ports.

We conclude this discussion of robustness by considering two additional pre-1956 infrastructure investments plausibly correlated with port depth. The first such infrastructure is naval bases. In the US, large military installations may promote local economic activity. If growth-yielding federal investments were concentrated near very deep ports, this could bias the coefficient on proximity to containerization upward. When we re-estimate Equation (1) using instrumental variables, omitting counties within 300 km of any naval base, coefficients are slightly larger and statistically indistinguishable from the main specification.[27]

Similarly, if very deep ports were crucial for oil importation, and oil importation caused population growth, our estimate of $\beta_1$ would be biased upward. A number

[27]As of the 1950s, the US had four domestic naval bases, at least 10 naval stations, and over 250 total facilities, which includes hospitals, test stations, air stations, and a large variety of other installations (U.S. Department of the Navy, 1952, 1959). Naval bases were Pearl Harbor, HI; San Diego, CA; Norfolk, VA and New London, CT. New London was actually taken out of "base" status between 1952 and 1959, but we include it for completeness. Relative to naval bases, naval stations are smaller, serve more limited purposes, and receive less investment (Coletta, 1985). Naval stations are so numerous that 300 km bands around them are indistinguishable from coastal locations; see our control for coastal locations in Table 3.

of factors argue against this interpretation. First, as of 1948, 90 percent of US oil was produced domestically and the US accounted for 62 percent of the world oil market (Mendershausen, 1950, p. 4). It was not until the 1970s, almost two decades after the advent of containerization, that the US was no longer able to fulfill oil demand with domestic oil.

Furthermore, port depth is not a key determinant for suitability as an oil port, allaying concerns about the validity of the instrument. During the period of domestic oil hegemony, most oil moved by pipeline, rather than by ship. Even when oil importation grew, port depth was not as crucial, because oil ships connect to offload via a pipeline, which can be quite long. Therefore, ships need not dock directly at the harbor to offload oil. Further, until the Suez Canal was dredged in the mid-1960s, it did not allow vessels with a draft deeper than 37 feet (Horn et al., 2010, p. 43).

Our analysis of robustness concludes by turning to a dataset of world ports and world cities to assess containerization's global impact. We focus primarily on the United States in this paper because of the rich data available at a relatively small geographic scale. However, containerization is clearly a global phenomenon, and one that may have had an even larger impact on economic activity in countries other than the United States. We use world population and port data to estimate regressions that parallel our main US regressions. We report results in Table 4. Columns 4 reports OLS results controlling for country fixed effects, the number of ports in 1953 within 300 km of each city (in 100 km distance bins), distance to the ocean, and log population in 1950. We find that cities within 100 km of a containerized port experience a 9 percentage point increase in population growth between 1950 and 2010 relative to cities more than 300 km away from a containerized port.

Just as in the US sample, we are concerned that the assignment of containerized ports to cities is not random, generating bias. Using the same instrumenting technique as in the US sample, we find that, similar to the US, proximity to a very deep port in 1953 is strongly related to proximity to a containerized port in 2010 (Appendix Table 3 presents summary statistics and Appendix Table 4 shows a strong first stage). The instrumental variable coefficients have the same signs as the OLS results, but are substantially larger. In the most complete specification in column 8, we find that cities within 200 km of a containerized port grow by an additional 30 percentage points, or about 20 percent of the mean. For cities between 200 and 300 km of a containerized port, we estimate a statistically insignificant increase in population growth of about 11 percentage points. These results are smaller in absolute terms than for the US, likely because we consider a sample of international cities that are relatively larger than the majority of US cities.

**Containerization's Impact on Other Economic Outcomes**

Having shown that proximity to a containerized port causes population growth, we test whether proximity to a containerized port also causes an increase in employment, nominal wages, industrial composition, and income. Using instrumental variables estimation with the full set of covariates from Table 2, column 1 in Table 5 confirms that, from 1956 to 2011, employment increases more in counties near containerized ports.[28] While only the coefficient for counties closest to a container port is statistically significant, the magnitude and pattern of employment increases is strikingly similar to what we find using Decennial Census population data. However, in comparison to the mean, these figures are substantially smaller. The mean change in log employment over the period is 1.13 (see final row), compared to a mean increase in log population of 0.37 (see final row in Table 3).

The dependent variable in Column 2 is nominal first quarter payroll per employee. Proximity to a containerized port is virtually unrelated to nominal payroll per employee.

[28] Employment data is from County Business Patterns (see details in Section 4).

As discussed in Section 3, the net effect on nominal wages is theoretically ambiguous because the home market effect and the market crowding effect go in opposite directions.

The middle two columns of Table 5 assess whether containerization changed the industrial composition of counties near containerized ports. Column 3 reports the share of employment in manufacturing, the industry most likely to produce products that travel in shipping containers. On average, across all counties, the share of employment in manufacturing declined by about 21 percentage points from 1956 to 2011 (last row). The coefficients reveal very little evidence of a smaller decline in manufacturing among treated counties.

Nonetheless, a more narrow focus on transportation does show relative growth. In column 4, the dependent variable is the share of employment in transportation services, which is "services which support transportation," and which includes "air traffic control services, marine cargo handling, and motor vehicle towing".[29] Relative to the miniscule one-tenth of one percent of employment in this industry on average, counties within 100 km of a container port see a statistically significant gain of three times this mean. Counties more than 100 km away from a container port see no significant change in this sector. Our finding that employment shifts towards transportation services is reminiscent of Michaels (2008) who finds that counties connected with highways experience an increase in trade-related activities, such as trucking and retail sales.

Finally, in the last three columns, we look at the impact of containerization on the income distribution. We look at income for the 10th, 50th, and 90th percentiles and find that counties within 100 km of a containerized port experience larger and significant increases in income across the whole distribution. In addition, as with population and overall employment, the pattern of decline with distance to the closest containerized port

[29]For 1956, we use SIC 47 for "services incidental to transportation," and for 2011 we use NAICS 488 for "support activities for transportation."

remains: counties farther away from a containerized port experience smaller additional increases in income relative to counties more than 300 km away from a containerized port.

## 6.3 Where Gains to Containerization Are Largest

In the previous subsections, we show that, on average, proximity to a containerized port causes increases in population and employment. We hypothesize that gains should be greater in initially low land value areas, and this section reports results from testing this claim.

We use three proxies for land values circa 1956. The first is the share of county employment in the manufacturing sector in 1956. Manufacturing was the high tech of the 1950s, and we anticipate that productive places should also be high land value places (Moretti, 2011). The second proxy is county population density as of 1950, and the third is assessed land value from the 1956 Census of Governments. While this last measure is the closest to a direct measure of the variable of interest, assessed values are notoriously different from market values. Particularly in this period, it was not unusual for assessment practices to vary substantially – and systematically – across jurisdictions (Anderson and Pape, 2010).

Table 6 reports coefficients on the measure of proximity to a containerized port and coefficients on the interaction of being below the median of variable $h_i$ and near a containerized port. Again, the dependent variable is the change in log population. The first column shows that half of the containerization-induced population growth in counties within 100 km of containerized ports occurs in counties with lower than median share of workers in manufacturing. For counties slightly farther from the port, almost all of the containerization-induced population growth occurs in counties with lower than median share of workers in manufacturing in 1956. While no initial condition explains as much of containerization-induced population growth as an initially small manufacturing sector, containerization-induced population growth is also large in initially less dense places (column 2). We observe no particular pattern in counties with low 1956 assessed land values (column 3).

Overall, these results paint a picture of containerization exerting the greatest influence not in dominant agglomerations—large, wealthy urban areas—but in second-tier agglomerations. These second-tier agglomerations are initially less dense and less concentrated in the vanguard technology of the 1950s (manufacturing). This is consistent with containerization's demand for large areas of land and suggests that containerization is easier to implement where land values are initially low.

These results are also consistent with a complementary story about the role of market access (e.g. Donaldson and Hornbeck, 2016). This line of argument says that containerization's impact will be larger, in percentage terms, in areas with initially low market access. This hypothesis is consistent with the results in column 4, showing larger gains in counties at the bottom half of highway intensity (highways per square km). We see no preferential pattern, however, with railroads (column 5).

# 7 Conclusion

Containerization is a fundamental engine of the global economy. Containerization simplifies and speeds packing, transit, pricing, and every transfer from ship to train to truck. It eliminates previously profitable pilferage and makes shipping more reliable. Since the advent of containerization in 1956, the cost of moving containerizable goods has plummeted.

In this paper, we analyze how local economic activity responds to the dramatic decline in trade costs brought by containerization. We use a novel cost-shifter instrument based on the historical depth of ports to show that, consistent with the predictions of a New Economic Geography model, containerization caused substantial population and employment growth in counties near container ports. These gains follow the pattern of decline with distance predicted by theory: counties closer to a containerized port experience larger increases than counties located farther away. Finally, consistent with containerization's need for substantial land for large cranes and vast marshalling yards, gains are located predominantly in counties with initially low population density and initially low manufacturing employment.

Whether and how containerization impacts the location of population, employment, and wages has implications for both the agglomerative forces that drive innovation, and for political representation that yields democratic outcomes. For policymakers to mitigate the uneven impacts of globalization, it is useful to first understand its causes.

## References


Anderson, Nathan B. and Pape, Andreas D., 2010. "A Reassessment of the 1970s Property Tax Revolt." Tech. Rep. Working Paper, University of Illinois at Chicago.

Angrist, Joshua, Lavy, Victor, and Schlosser, Analia, 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." Journal of Labor Economics 28(4): 773–824.

Autor, David H., Dorn, David, and Hanson, Gordon H., 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." American Economic Review 103(6): 2121–68.

Baum-Snow, Nathaniel, 2007. "Did Highways Cause Suburbanization?" Quarterly Journal of Economics 122(2): 775–805.

Bernhofen, Daniel, El-Sahli, Zouheir, and Kneller, Richard, 2016. "Estimating the Effects of the Container Revolution on World Trade." Journal of International Economics 98: 36–50.

Bridgman, Benjamin, 2014. "Why Containerization Did Not Reduce Ocean Trade Shipping Costs." Tech. rep. Working Paper.

Broeze, Frank, 2002. The Globalisation of the Oceans: Containerisation From the 1950s to the Present, vol. 23 of Research in Maritime History. St. John's, Newfoundland: International Maritime Economic History Association.

Campante, Filipe and Yanagizawa-Drott, David, 2017. "Long-Range Growth: Economic Development in the Global Network of Air Links*." The Quarterly Journal of Economics page qjx050.

Chinitz, Benjamin, 1960. Freight and the Metropolis: The Impact of America's Transport Revolutions on the New York Region. Havard University Press.

Coşar, Kerem and Demir, Banu, 2018. "Shipping Inside the Box: Containerization and Trade." Tech. Rep. Discussion Paper No. DP11750, CEPR.

Coletta, Paolo E., 1985. "Preface." In Paolo E. Coletta and Jack K. Bauer, (Eds.) "United States Navy and Marine Corps Bases, Domestic," Greenwood Press.

Donaldson, Dave and Hornbeck, Richard, 2016. "Railroads and American Economic Growth: A Market Access Approach." The Quarterly Journal of Economics 131(2): 799–858.

Duranton, Gilles, Morrow, Peter M., and Turner, Matthew A., 2014. "Roads and Trade: Evidence from the US." Review of Economic Studies 81: 681–724.

Duranton, Gilles and Turner, Matthew, 2012. “Urban Growth and Transportation.” Review of Economic Studies 79: 1407–1440.

Feyrer, James, 2009a. “Distance, Trade, and Income - The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment.” NBER Working Papers 15557, National Bureau of Economic Research, Inc.

Feyrer, James, 2009b. “Trade and Income – Exploiting Time Series in Geography.” NBER Working Papers 14910, National Bureau of Economic Research, Inc.

Helpman, E., 1995. “The Size of Regions.” Papers 14-95, Tel Aviv.

Holmes, Thomas J. and Singer, Ethan, 2017. “Indivisibilities in Distribution.” Tech. Rep. Working Paper 739, Federal Reserve Bank of Minneapolis.

Horn, Kevin, Wilson, Erin, Knight, Kevin, and Durden, Susan, 2010. “National Economic Development (NED) Manual For Deep Draft Navigation.” Tech. rep., U. S. Army Engineer Institute for Water Resources.

Hummels, David, 2007. “Transportation Costs and International Trade in the Second Era of Globalization.” Journal of Economic Perspectives 21(3): 131–154.

Hummels, David and Schaur, Georg, 2013. “Time as a Trade Barrier.” American Economic Review 103(7): 2935–2959.

Kendall, L.C., 1986. The Business of Shipping. Cornell Maritime Press, fifth edn.

Kovak, Brian K., 2013. “Regional Effects of Trade Reform: What Is the Correct Measure of Liberalization?” American Economic Review 103(5): 1960–1976.

Kuby, Michael and Reid, Neil, 1992. “Technological Change and the Concentration of the U.S. General Cargo Port System: 1970-88.” Economic Geography 68(3): 272–289.

Levinson, Mark, 2008. The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger. Princeton, NJ: Princeton University Press.

Mendershausen, Horst, 1950. “Dollar Shortage and Oil Surplus in 1949-1950.” Tech. Rep. No. 11, International Finance Section, Department of Economics and Social Institutions, Princetion Univeristy.

Michaels, Guy, 2008. “The effect of trade on the demand for skill: Evidence from the Interstate Highway System.” Review of Economics and Statistics 09: 683–701.

Moretti, Enrico, 2004. “Human Capital Externalities in Cities.” In Vernon Henderson and Jacques Thisse, (Eds.) “Handbook of Regional and Urban Economics,” Elsevier.

Moretti, Enrico, 2011. “Local Labor Markets.” In David Card and Orley Ashenfelter, (Eds.) “Handbook of Labor Economics,” vol. 4, pages 1237 – 1313. Elsevier.

Morgan, F. W., 1952. Ports and Harbours. London, UK: Hutchison's University Library.

National Geospatial-Intelligence Agency, 1953. "World Port Index, 1953." H.O. Pub. No. 950, First Edition.

National Geospatial-Intelligence Agency, 2015. "World Port Index, 2015." Twenty-Fourth Edition.

Pascali, Luigi, 2017. "The Wind of Change: Maritime Technology, Trade, and Economic Development." American Economic Review 107(9): 2821–54.

Rappaport, Jordan and Sachs, Jeffrey D, 2003. "The United States as a Coastal Nation." Journal of Economic Growth 8(1): 5–46.

Redding, Stephen and Sturm, Daniel M., 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." American Economic Review 98(5): 1766–1797.

Redding, Stephen J. and Rossi-Hansberg, Esteban, 2017. "Quantitative Spatial Economics." Annual Review of Economics 9(1): 21–58.

Redding, Stephen J. and Turner, Matthew A., 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In Gilles Duranton and William Strange, (Eds.) "Handbook of Regional and Urban Economics," Elsevier B. V.

Rodrigue, Jean-Paul, 2017. The Geography of Transport Systems. Hofstra University, Department of Global Studies and Geography.

Romer, David H. and Frankel, Jeffrey A., 1999. "Does Trade Cause Growth?" American Economic Review 89(3): 379–399.

Rua, Gisela, 2014. "Diffusion of Containerization." Finance and Economics Discussion Series 88, Board of Governors of the Federal Reserve System (U.S.).

Sanderson, Eleanor and Windmeijer, Frank, 2016. "A weak instrument F-test in linear IV models with multiple endogenous variables." Journal of Econometrics 190(2): 212–221.

Sargent, A. J., 1938. Seaports and Hinterlands. London, UK: Adam and Charles Black.

Topalova, Petia, 2010. "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India." American Economic Journal: Applied Economics 2(4): 1–41.

United Nations Conference on Trade and Development, 2013. Review of Maritime Transport. New York, NY: United Nations.

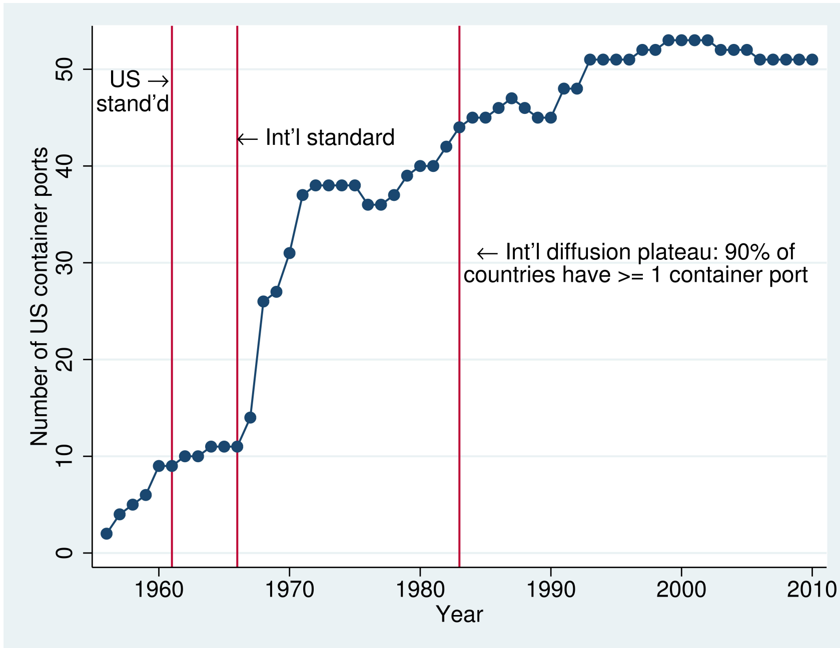
U.S. Department of the Navy, 1952. "Catalog of Naval Shore Activities." Tech. Rep. OPNAV P213-105.

U.S. Department of the Navy, 1959. "Catalog of Naval Shore Activities." Tech. Rep. OPNAV P09B23-105.

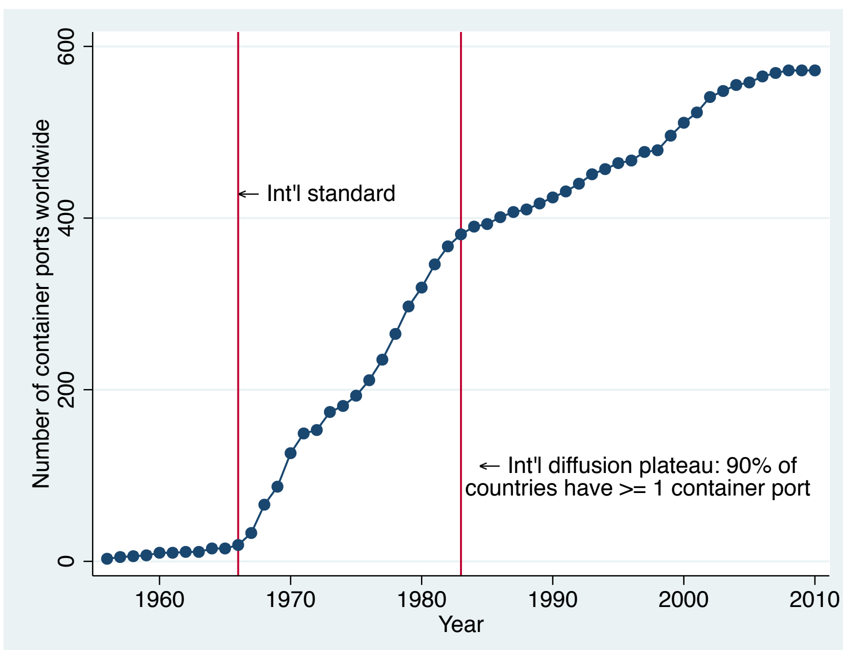
Wilson, R. P., 1982. "The Containerization of Ports." UC Santa Cruz Working Paper.

Figure 1: Adoption of Containerization: 1956–2008

(a) United States



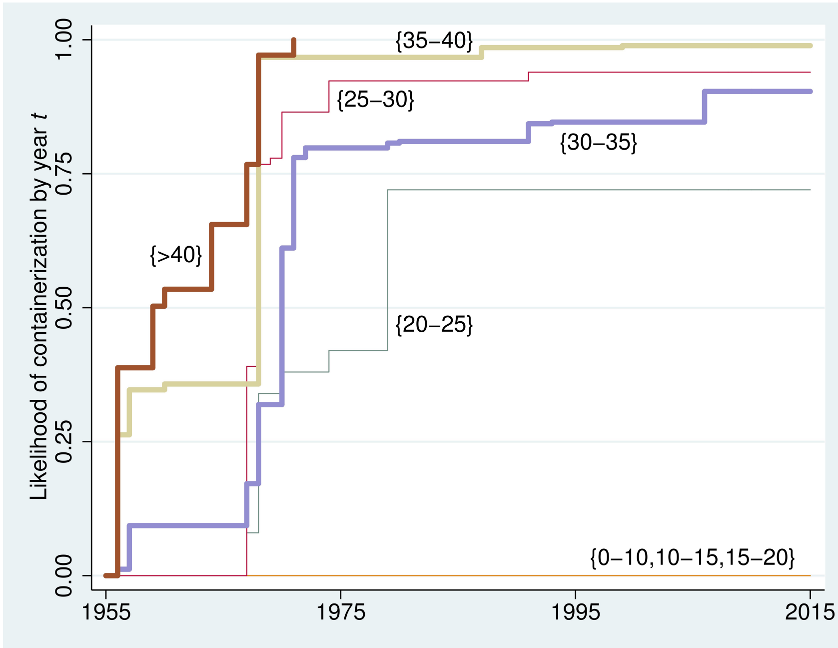

(b) Worldwide



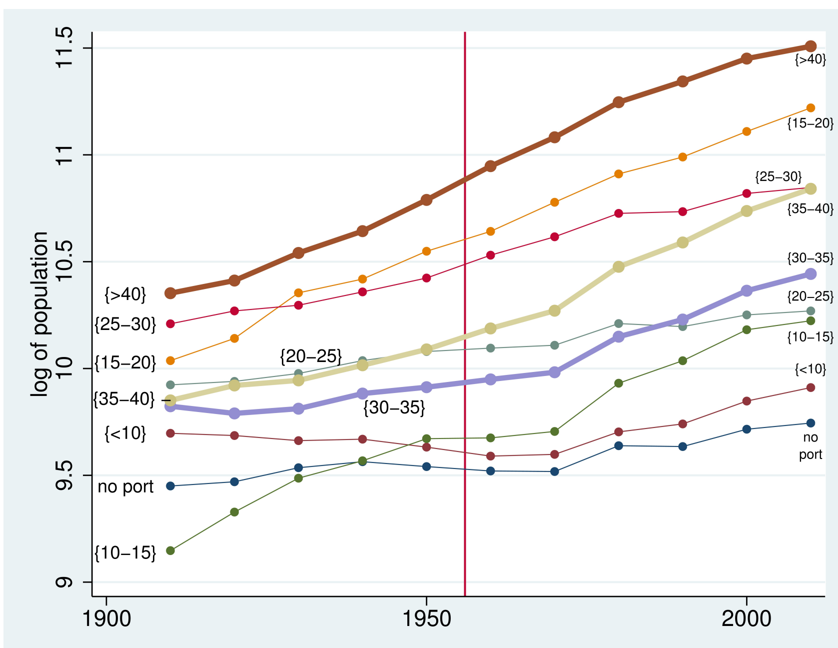

Note: The upper panel shows the diffusion of containerization across US ports; the bottom panel repeats this exercise for world ports. Source: *Containerisation International Yearbook*, volumes 1968 and 1970–2010.

Figure 2: Graphical Intuition

(a) First Stage: Depth and Likelihood of Containerization



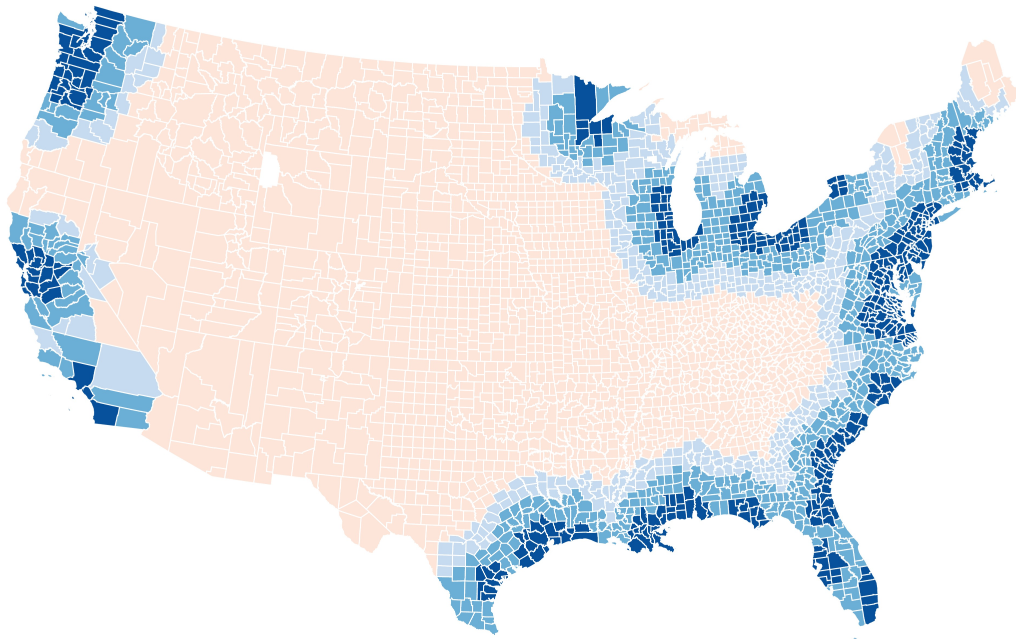

(b) Reduced Form: Depth and Population Changes



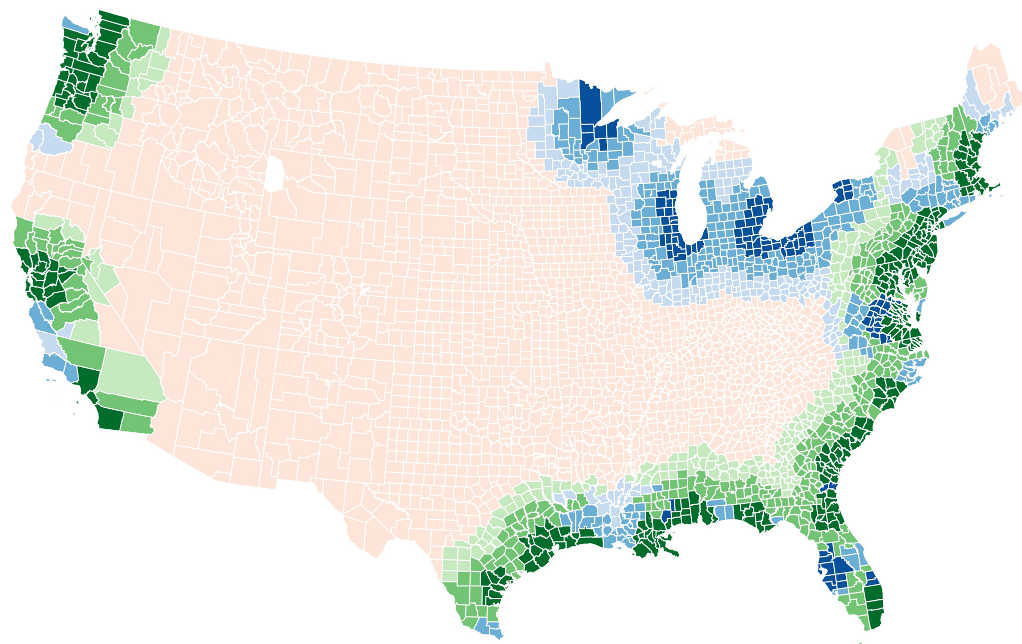

Notes: In both figures, thick lines denote depths that we label "very deep" in our estimation. Figure 2a shows the likelihood that a county will have a containerized port within 300 km in year $t$ by the depth of the deepest port within 300 km in 1953. On average, deeper ports are more likely to ever containerize, and more likely to containerize early. Figure 2b plots the logarithm of population over time by the depth of the deepest port within 300 km in 1953.

Figure 3: Geographic Variation in Treatment and Instrument

(a) Counties Near a Containerized Port in 2010

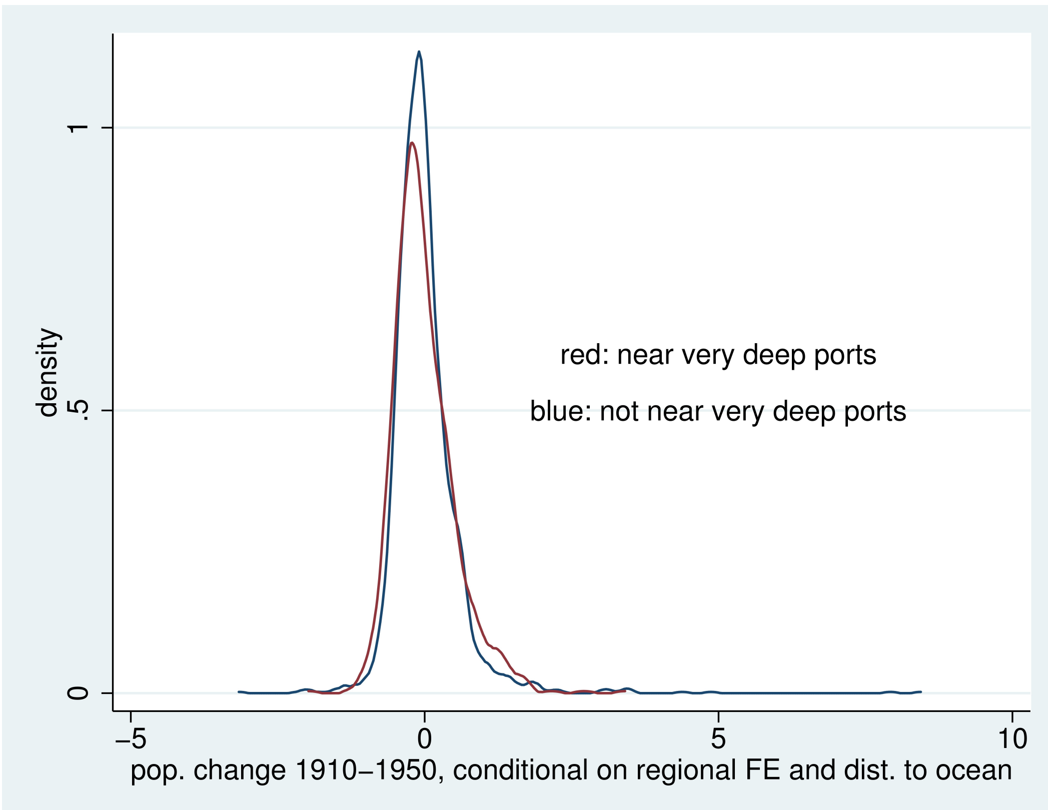


(b) Counties Near a Containerized Port in 2010 and Near a Very Deep Port in 1953



Notes: Figure 3a shows the distance to the nearest containerized port in 2010. Blue polygons are counties $d_1$ to $d_2$ km from the nearest containerized port. Distance bins $\{d_1, d_2\}$ are {0 to 100, 100 to 200, 200 to 300}. Figure 3b shows the distance to the nearest containerized port in 2010 as well as the distance to the nearest "very deep" port in 1953. Green colors represent counties that are $d_1$ to $d_2$ km from the nearest containerized port and $d_1$ to $d_2$ km from the nearest "very deep" port in 1953.

Figure 4: Port Depth Unrelated to Pre-Containerization Growth




Notes: This picture shows the distribution of county population change 1910 to 1950, conditional on regional fixed effects and distance to the ocean. Counties near very deep ports are in red and those not near very deep ports are in blue. Regressions results show no significant difference between these two means.

Table 1: County Characteristics by Distance to Nearest Containerized Port

| | Distance to Containerized Port, km | | | | |
|---|---|---|---|---|---|
| | 0 to 100 | 100 to 200 | 200 to 300 | Ever Cont. | Never Cont. |
| | (1) | (2) | (3) | (4) | (5) |
| Log Population | | | | | |
| 1910 | 10.31 | 10.03 | 10.02 | 10.11 | 9.47 |
| | [ 1.22] | [ 0.82] | [ 0.80] | [ 0.95] | [ 0.96] |
| 1950 | 10.81 | 10.23 | 10.14 | 10.36 | 9.58 |
| | [ 1.47] | [ 0.97] | [ 0.97] | [ 1.16] | [ 0.96] |
| 2010 | 11.70 | 10.75 | 10.52 | 10.94 | 9.79 |
| | [ 1.50] | [ 1.16] | [ 1.15] | [ 1.35] | [ 1.32] |
| Log Employment | | | | | |
| 1956 | 9.02 | 8.19 | 8.04 | 8.37 | 7.18 |
| | [ 1.94] | [ 1.44] | [ 1.45] | [ 1.65] | [ 1.43] |
| 2011 | 10.37 | 9.31 | 9.08 | 9.53 | 8.35 |
| | [ 1.83] | [ 1.45] | [ 1.47] | [ 1.66] | [ 1.55] |
| Log Payroll Per Employee | | | | | |
| 1956 | -0.27 | -0.37 | -0.40 | -0.35 | -0.50 |
| | [ 0.33] | [ 0.29] | [ 0.31] | [ 0.31] | [ 0.32] |
| 2011 | 2.19 | 2.04 | 2.02 | 2.08 | 1.97 |
| | [ 0.29] | [ 0.20] | [ 0.19] | [ 0.24] | [ 0.22] |
| Region | | | | | |
| Northeast | 0.19 | 0.17 | 0.12 | 0.16 | 0.00 |
| Midwest | 0.19 | 0.28 | 0.38 | 0.29 | 0.39 |
| South | 0.49 | 0.48 | 0.45 | 0.47 | 0.43 |
| West | 0.13 | 0.07 | 0.05 | 0.08 | 0.17 |
| Share Employment, Manufacturing | | | | | |
| 1956 | 0.42 | 0.41 | 0.42 | 0.42 | 0.26 |
| | [ 0.19] | [ 0.19] | [ 0.20] | [ 0.19] | [ 0.22] |
| 2011 | 0.10 | 0.15 | 0.14 | 0.13 | 0.10 |
| | [ 0.09] | [ 0.12] | [ 0.12] | [ 0.11] | [ 0.12] |
| Observations | 370 | 523 | 442 | 1335 | 1688 |

Note: This table reports means and standard deviations in brackets. The number of observations at the bottom of the table applies to all variables except the 1910 population and the payroll and employment variables; each has slightly fewer observations.

Table 2: Containerization Associated with Increased Population, Particularly Near the Port

| | OLS | | | | IV | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Closest container port is | | | | | | | | |
| 0 to 100 km | 0.684*** | 0.604*** | 0.464*** | 0.453*** | 0.685*** | 0.642*** | 0.410** | 0.529*** |
| | (0.064) | (0.063) | (0.095) | (0.094) | (0.082) | (0.087) | (0.187) | (0.18) |
| 100 to 200 km | 0.348*** | 0.351*** | 0.256*** | 0.249*** | 0.237*** | 0.371*** | 0.219 | 0.285* |
| | (0.054) | (0.053) | (0.078) | (0.076) | (0.086) | (0.087) | (0.154) | (0.147) |
| 200 to 300 km | 0.235*** | 0.202*** | 0.156** | 0.132* | 0.215** | 0.267*** | 0.175 | 0.204 |
| | (0.057) | (0.056) | (0.07) | (0.07) | (0.097) | (0.102) | (0.14) | (0.139) |
| Covariates | | | | | | | | |
| Regional fixed effects | x | x | x | x | x | x | x | x |
| Demographics | | x | x | x | | x | x | x |
| Log of population, 1920-1940 | | x | x | x | | x | x | x |
| Distance to the ocean | | | x | x | | | x | x |
| Number of 1953 ports | | | x | x | | | x | x |
| Total int'l trade at ports, 1955 | | | x | x | | | x | x |
| 1950s-era transportation | | | | x | | | | x |
| Share manufacturing employment, 1956 | | | | x | | | | x |
| R-squared | 0.186 | 0.328 | 0.356 | 0.372 | 0.183 | 0.327 | 0.355 | 0.371 |
| Kleinberg-Paap F Stat | | | | | 99.1 | 95.7 | 21 | 21.1 |

Notes: Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. All regressions use 3,023 observations and cluster standard errors at the 2010 commuting zone. The dependent variable is the change in log population, 1950-2010. The mean of the dependent variable is 0.373. Demographics is share of people with a college degree or more and share African America, both measured as of 1950. Number of 1953 ports and total international trade at ports in 1955 are both vectors with totals by 100 km bins. 1950s-era transportation is a vector which measures the kilometers of highways c. 1960, kilometers of navigable waterways, and kilometers of railroads c. 1957 in each county, all per square kilometer of land area. See data appendix for complete details on years and sources.

Table 3: Impact of Containerization Robust to Alternative Specifications

| | Main spec., Table 2, col. 8 | With R & S coast control | Within 400 km of a 1953 port | Squares of population | Omit Midwest region | Additional demographic covariates |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Closest Container Port is within | | | | | | |
| 0 to 100 km | 0.529*** | 0.423** | 0.510** | 0.448** | 0.443** | 0.449** |
| | (0.18) | (0.201) | (0.202) | (0.182) | (0.201) | (0.183) |
| 100 to 200 km | 0.285* | 0.260* | 0.237 | 0.231 | 0.174 | 0.236 |
| | (0.147) | (0.146) | (0.176) | (0.147) | (0.166) | (0.145) |
| 200 to 300 km | 0.204 | 0.205 | 0.228 | 0.171 | 0.109 | 0.164 |
| | (0.139) | (0.139) | (0.173) | (0.139) | (0.146) | (0.136) |
| R-squared | 0.357 | 0.363 | 0.311 | 0.371 | 0.293 | 0.371 |
| Kleinberg-Paap F Stat | 21.3 | 17.7 | 21.4 | 21.6 | 24.6 | 21.3 |
| Observations | 3023 | 3023 | 1767 | 3023 | 1975 | 3023 |
| Mean, dependent variable | 0.373 | 0.373 | 0.514 | 0.373 | 0.508 | 0.373 |

Notes: Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. All specifications instrumental variable regressions with clustered standard errors at the 2010 commuting zone. Log of population is the dependent variable and all regressions include the most complete covariate list from Table 2. Column 1 repeats the most saturated estimation from Table 2 Column 8. Column 2 controls for the Rappaport and Sachs (2003) measure of coastal proximity. Column 3 restricts the sample to counties within 400 km of a 1953 port. Column 4 includes squares of 1920, 1930 and 1940 population. Column 5 omits the Midwest census region, which has no very deep ports. Column 6 includes additional demographic covariates measured in 1950: share of people 25 or older with less than a high school degree, share foreign born, government workers per capita, and share age 65 and older.

Table 4: Containerization Impacts Growth in World Cities

| | OLS | | | | IV | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Closest container port is | | | | | | | | |
| 0 to 100 km | -0.010 | 0.069 | 0.007 | 0.090 | 0.047 | 0.134* | 0.216* | 0.310*** |
| | (0.056) | (0.056) | (0.066) | (0.065) | (0.071) | (0.070) | (0.122) | (0.120) |
| 100 to 200 km | -0.040 | -0.035 | -0.037 | -0.015 | 0.188** | 0.165* | 0.310** | 0.307*** |
| | (0.060) | (0.058) | (0.067) | (0.066) | (0.091) | (0.086) | (0.124) | (0.118) |
| 200 to 300 km | -0.038 | -0.027 | -0.035 | -0.012 | 0.040 | 0.027 | 0.114 | 0.113 |
| | (0.064) | (0.060) | (0.067) | (0.064) | (0.112) | (0.105) | (0.127) | (0.120) |
| Covariates | | | | | | | | |
| Country fixed effects | x | x | x | x | x | x | x | x |
| Log of population, 1950 | | x | | x | | x | | x |
| Distance to the ocean | | | x | x | | | x | x |
| Number of 1953 ports | | | x | x | | | x | x |
| R-squared | 0.655 | 0.684 | 0.663 | 0.690 | 0.648 | 0.678 | 0.652 | 0.680 |
| Kleinberg-Paap F Stat | | | | | 43.7 | 43.9 | 42.5 | 43.6 |

Notes: Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. All regressions use 1,051 observations, and the unit of observation is a city with as least 50,000 inhabitants in 1950. The dependent variable is the change in log population, 1950 to 2010. The mean of the dependent variable is 1.54.

Table 5: More Employment and Higher Earnings Near Containerized Ports

| | IV, Dependent Variable is | | | | | | |
|---|---|---|---|---|---|---|---|
| | All industries | | Employment Share | | Log $p$th percentile income, where $p$ is | | |
| | Log employment | Log payroll/ employee | Manufacturing | Transportation Services | 10 | 50 | 90 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Closest container port is | | | | | | | |
| 0 to 100 km | 0.347* | 0.049 | -0.019 | 0.003** | 0.175* | 0.276*** | 0.152** |
| | (0.201) | (0.068) | (0.021) | (0.001) | (0.083) | (0.069) | (0.055) |
| 100 to 200 km | 0.124 | 0.043 | 0 | 0.001 | 0.082 | 0.109* | 0.084+ |
| | (0.156) | (0.044) | (0.017) | (0.001) | (0.067) | (0.055) | (0.045) |
| 200 to 300 km | 0.02 | 0.018 | 0.008 | 0 | 0.058 | 0.05 | -0.01 |
| | (0.147) | (0.039) | (0.017) | (0.001) | (0.067) | (0.056) | (0.046) |
| R-squared | 0.178 | 0.155 | 0.76 | 0.075 | 0.298 | 0.436 | 0.324 |
| Kleinberg-Paap F Stat | 21.1 | 22 | 21.1 | 21.1 | 21.4 | 21.4 | 21.4 |
| Observations | 2985 | 2981 | 2985 | 2985 | 3022 | 3022 | 3022 |
| Mean, Dependent Variable | 1.135 | 2.448 | -0.215 | 0.001 | 3.547 | 3.147 | 3.176 |

Notes: Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. All specifications are instrumental variable regressions with County Business Patterns and Census income data, and include the most complete covariate list from Table 2. We cluster the standard errors at the 2010 commuting zone. The second pair of columns report fewer observations because some counties are sufficiently small to suppress all payroll information.

Table 6: Greater Containerization-Induced Growth in Initially Lagging Places

| | 1950 Interaction Variable is | | | | |
|---|---|---|---|---|---|
| | Manuf. share of Employmt | 1950 Population Density | 1956 Assessed Land Value | Highway km / county sq km | Rail km / county sq km |
| | (1) | (2) | (3) | (4) | (5) |
| Closest container port is within | | | | | |
| 0 to 100 km | 0.293 | 0.293 | 0.559** | 0.094 | 0.375 |
| | (0.224) | (0.192) | (0.255) | (0.218) | (0.231) |
| 100 to 200 km | -0.072 | 0.16 | 0.521** | -0.01 | 0.293 |
| | (0.168) | (0.159) | (0.235) | (0.186) | (0.193) |
| 200 to 300 km | -0.105 | 0.056 | 0.231 | -0.038 | 0.127 |
| | (0.144) | (0.141) | (0.196) | (0.182) | (0.167) |
| Container port distance * 1{County $\leq$ median(column header variable)} | | | | | |
| 0 to 100 km | 0.267* | 0.389*** | 0.038 | 0.523*** | 0.166 |
| | (0.145) | (0.138) | (0.159) | (0.123) | (0.12) |
| 100 to 200 km | 0.531*** | 0.218** | -0.238 | 0.314** | -0.032 |
| | (0.141) | (0.1) | (0.145) | (0.131) | (0.123) |
| 200 to 300 km | 0.467*** | 0.282** | 0.004 | 0.267* | 0.104 |
| | (0.147) | (0.113) | (0.147) | (0.152) | (0.136) |
| R-squared | 0.368 | 0.368 | 0.355 | 0.377 | 0.358 |
| Kleinberg-Paap F Stat | 10.9 | 10.6 | 7.6 | 12.7 | 6.9 |
| Median, interaction variable | 0.44 | 17.1 | 0.01 | 0 | 0.07 |
| Share of observations $\leq$ median | | | | | |
| 0 to 100 km | 0.48 | 0.37 | 0.35 | 0.66 | 0.44 |
| 100 to 200 km | 0.49 | 0.54 | 0.52 | 0.81 | 0.54 |
| 200 to 300 km | 0.53 | 0.57 | 0.6 | 0.81 | 0.50 |

Note: Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. All specifications are instrumental variable estimates of Equation (3) with population as the dependent variable. All regressions have 3,023 observations and cluster standard errors at the 2010 commuting zone. The first panel of coefficients reports the average impact of containerization by distance from the port; the second panel of coefficients reports whether there is any additional population growth if the county's value of variable $h_i$ in 1950 is below the median among treated observations.

FOR ONLINE PUBLICATION

# A Data Appendix

## A.1 Data Sources

We use data from a variety of sources. This appendix provides source information.

1. County Business Patterns
   These data include total employment, total number of establishments (with some variation in this definition over time), and total payroll.

   - 1956: Courtesy of Gilles Duranton and Matthew Turner. See Duranton et al. (2014) for source details. We collected a small number of additional counties that were missing from the Duranton and Turner data.
     - In these data, payroll is defined as the "amount of taxable wages paid for covered employment [covered by OASI, or almost all "nonfarm industrial and commercial wage and salary employment" (page VII)[30]] during the quarter. Under the law in effect in 1956, taxable wages for covered employment were all payments up to the first $4,200 paid to any one employee by any one employer during the year, including the cash value of payments in kind. In general, all payments for covered employment in the first quarter were taxable unless the employee was paid at the rate of more than $16,800 per year. For the first quarter of 1956, it is estimated that 97.0 percent of total non-agricultural wages and salaries in covered employment was taxable. The taxable proportion of total wages becomes smaller in the later quarter of the year. Data are presented for the first quarter because wages for this quarter are least affected by the provisions of the law limiting taxable wages to $4,200 per year." (page VI, Section III, Definitions in 1956 County Business Patterns report.)
   - 1967 to 1985: U.S. National Archives, identifier 313576.
   - 1986 to 2011: U.S. Census Bureau. Downloaded from `https://www.census.gov/econ/cbp/download/`
     - For comparability, we also use total first quarter payroll from these data.

2. Decennial Census: Population and demographics data by county

   - 1910: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)

[30]Data also exclude railroad employment.

- 1920: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)
- 1930: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)
- 1940: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)
- 1950
  - ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1950 Census I (County and State)
  - Census of Population, 1950 Volume II, Part I, Table 32.
- 1960: ICPSR 02896, Historical, Demographic, Economic and Social Data: The United States, 1790-2002, Dataset 38: 1960 Census I (County and State)
- 1970: ICPSR 8107, Census of Population and Housing, 1970: Summary Statistic File 4C – Population [Fourth Count]
- 1980: ICPSR 8071, Census of Population and Housing, 1980: Summary Tape File 3A
- 1990: ICPSR 9782, Census of Population and Housing, 1990: Summary Tape File 3A
- 2000: ICPSR 13342, Census of Population and Housing, 2000: Summary File 3
- 2010: U.S. Census Bureau, 2010 Decennial Census Summary File 1, Downloaded from `http://www2.census.gov/census_2010/04-Summary_File_1/`
- 2010 (2008-2012): U.S. Census Bureau, American Community Survey, 5-Year Summary File, downloaded from `http://www2.census.gov/acs2012_5yr/summaryfile/2008-2012_ACSSF_All_In_2_Giant_Files%28Experienced-Users-Only%29/`

3. Port Universe and Depth
   - We use these documents to establish the population of ports in any given year.
   - 1953: *World Port Index*, National Geospatial-Intelligence Agency (1953)
   - 2015: *World Port Index*, National Geospatial-Intelligence Agency (2015)
4. Port Containerization Adoption Year
   - 1956–2010: *Containerisation International Yearbook* for 1968 and 1970–2010
5. Port Volume: Total imports and exports by port
   - 1948: United States Foreign Trade, January-December 1949: Water-borne Trade by United States Port, 1949, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census. FT 972.

- 1955: United States Waterborne Foreign Trade, 1955, Washington, D.C. : U.S. Dept. of Commerce, Bureau of the Census. FT 985.
- 2008: *Containerisation International* yearbook 2010, pp. 8–11.

6. Highways

   - 2014: 2014 National Transportation Atlas, Office of the Assistant Secretary for Research and Technology, Bureau of Transportation Statistics, United States Department of Transportation. `http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2014/index.html`.
   - c. 1960: Office of Planning, Bureau of Public Roads, US Department of Commerce, "The National System of Interstate and Defense Highways." Library of Congress Call number G3701.P21 1960.U5. Map reports improvement status as of December 31, 1960.

7. Railways

   - 2014: 2014 National Transportation Atlas, Office of the Assistant Secretary for Research and Technology, Bureau of Transportation Statistics, United States Department of Transportation. `http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2014/index.html`.
   - c. 1957: Army Map Service, Corps of Engineers, US Army, "Railroad Map of the United States," prepared 1935, revised April 1947 by AMS. 8204 Edition 5-AMS. Library of Congress call number G3701.P3 1957.U48.

8. Waterways

   - 2014: 2014 National Transportation Atlas, Office of the Assistant Secretary for Research and Technology, Bureau of Transportation Statistics, United States Department of Transportation. `http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2014/index.html`.

9. World Population Data: World Urbanization Prospects, 2014 Revision

   - Population counts for all urban agglomerations whose populations exceed 300,000 at any time between 1950 and 2010.
   - Produced by the United Nations, Department of Economic and Social Affairs, Population Division.
   - Downloaded from `http://esa.un.org/unpd/wup/CD-ROM/WUP2014_XLS_CD_FILES/WUP2014-F22-Cities_Over_300K_Annual.xls`

10. Property value data

    - 1956: 1957 Census of Governments: Volume 5, *Taxable Property Values in the United States*
    - 1991: 1992 Census of Governments, Volume 2 *Taxable Property Values*, Number 1 *Assessed Valuations for Local General Property Taxation*
    - In both 1957 and 1992, the Census reports a total figure for the New York City, which consists of five separate counties (equivalent to the boroughs). We attribute the total assessed value from the census of governments to each county (borough) by using each borough's share of total assessed value. For 1956, we rely upon the *Annual Report of the Tax Commission and the Tax Department to the Mayor of the City of New York* as of June 30, 1956, page 23 which reports "Assessed Value of All Real Estate in New York City for 1956-1957." For 1991, we rely upon *Department of Finance Annual Report, 1991-1992*, pages 19-24.
    - The District of Columbia is missing an assessed value for 1956 in the Census of Government publication listed above. However, the amount is available in *Trends in Assessed Valuations and Sales Ratios, 1956-1966*, US Department of Commerce, Bureau of the Census, March 1970. We use this value.
    - For 2010 value, we use the sum of the value of aggregate owner occupied stock (American Community Survey) and the aggregate value of the rental occupied stock. As the Census only reports aggregate gross rent, we convert aggregate gross rent to aggregate value of the rental stock by multiplying the aggregate value of the rental stock (by 12 to generate a monthly figure) by the average rent-price ratio for years 2008-2012 (corresponding with the ACS years) from Lincoln Institute Rent-price ratio data[31].

## A.2 Data Choices

1. U.S. County Sample

   Our unit of analysis is a consistent-border county from 1950 to 2010. We generate these counties by aggregating 1950 counties. Please see the final Appendix Table for the specific details of aggregation.[32]

   The 1956 County Business Patterns allowed for reporting of only 100 jurisdictions per state, leading to the reporting of aggregate values for agglomerations of counties in states with many counties. See Duranton et al. (2014) for the initial collection of these data, and additional details. To resolve the problem of making these 1956

[31] http://datatoolkits.lincolninst.edu/subcenters/land-values/rent-price-ratio.asp

[32] These groupings relied heavily on the very helpful work of the Applied Population Laboratory group at the University of Wisconsin. See their documentation at http://www.netmigration.wisc.edu/datadictionary.pdf.

units consistent with the 1950 census units, we disaggregate the 1956 CBP data in the agglomerated reporting into individual counties, attributing economic activity by population weights.

Alaska and Hawaii were not states in 1950. We omit Alaska from our sample, because in 1950 it has only judicial districts, which do not correspond to modern counties. We keep Hawaii, where the 1950 borders are relatively equivalent to modern counties. We also keep Washington, DC, in all years.

We also make a few additional deletions

- Two counties that only appear in the data (1910-1930) before our major period of analysis: Campbell, GA (13/041) and Milton, GA (13/203).
- Two problematic counties. Menominee, WI (55/078) created in 1959 out of an Indian reservation; it has very few people. Broomfield, CO (08/014), created in 2001 from parts of four other counties.
- Two counties where land area changes are greater than 40 percent. These are Denver County, CO (08/031) and Teton County, WY (56/039).

2. County Business Patterns data

   - For some county/industry groupings, there is a disclosure risk in reporting either the total number of employees or the total payroll. In such cases, we convert the disclosure code ("D" in the years before 1974) to 0.
   - "Payroll" is first quarter payroll.

3. Income distribution calculations

   - We use binned income data. In 1950, the number in each bin is the total number of families and unrelated individuals. To be consistent, in 2010 we also use the total number of families plus unrelated individuals.
   - To calculate percentiles, we assume that income is uniformly distributed within bins, with the exception of the top bin, which has no top code.
   - For the top bin, we assume that income is distributed following a Pareto distribution, with a parameter $\alpha$. We assume that $\alpha = \max(\hat{\alpha}, 1)$. Let $N_B$ be the number of people in the top income bin, and $N_{B-1}$ be the number of people in the second highest bin. Similarly, $L_B$ be the lower bound of the top income bin and $L_{B-1}$ be the lower bound of the second highest income bin. Then

$$\hat{\alpha} = \frac{\log(N_B + N_{B-1}) - \log(N_B)}{log(L_B) - log(L_{B-1})}$$

.

Appendix Figure 1: Evolution of Ship Sizes

WWII technology

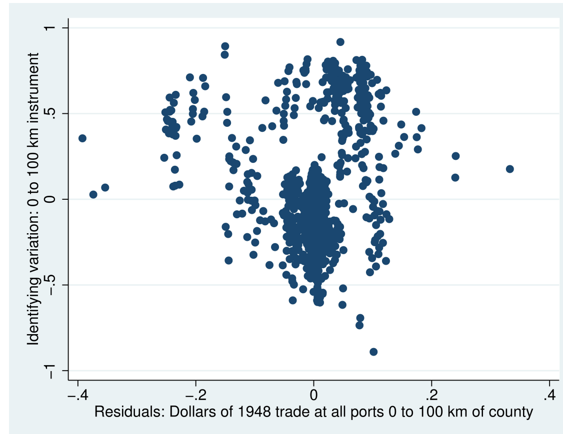

First container ships, 1956 to 1970s


Today, Post-Panamax

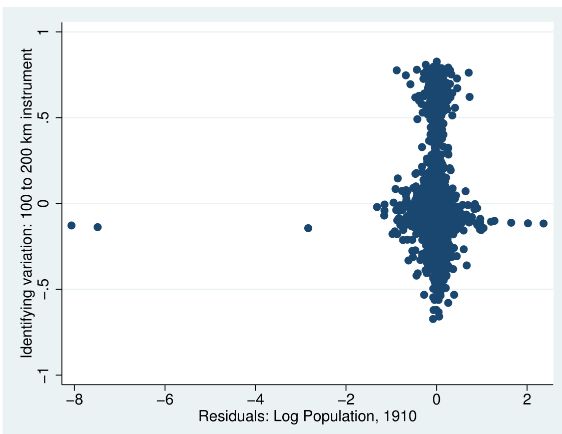

Source: WWII, authors; remaining ships, (Rodrigue, 2017).

Appendix Figure 2: Instrument Variation vs. Pre-Treatment Covariates: All Instruments

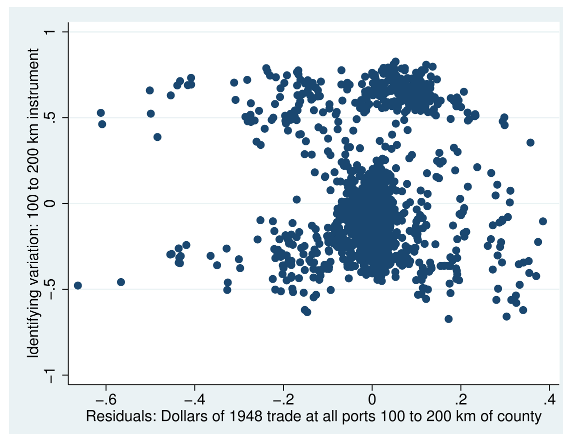

(a)



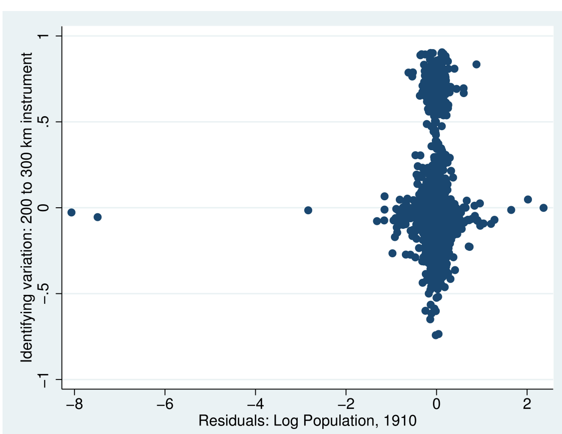

(b)



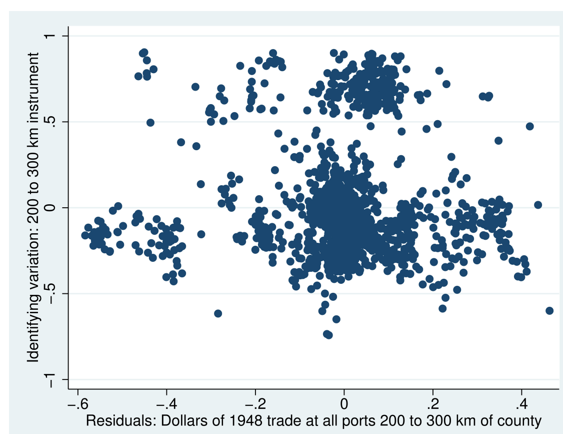

(c)




(d)



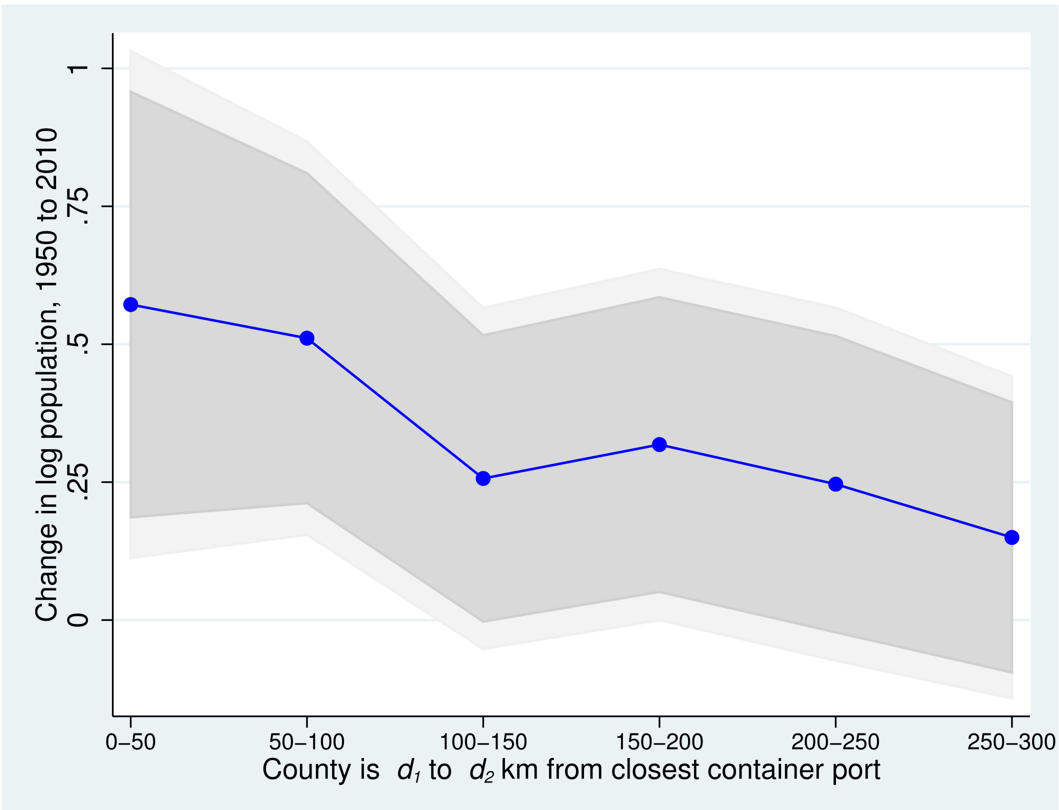

(e)




(f)

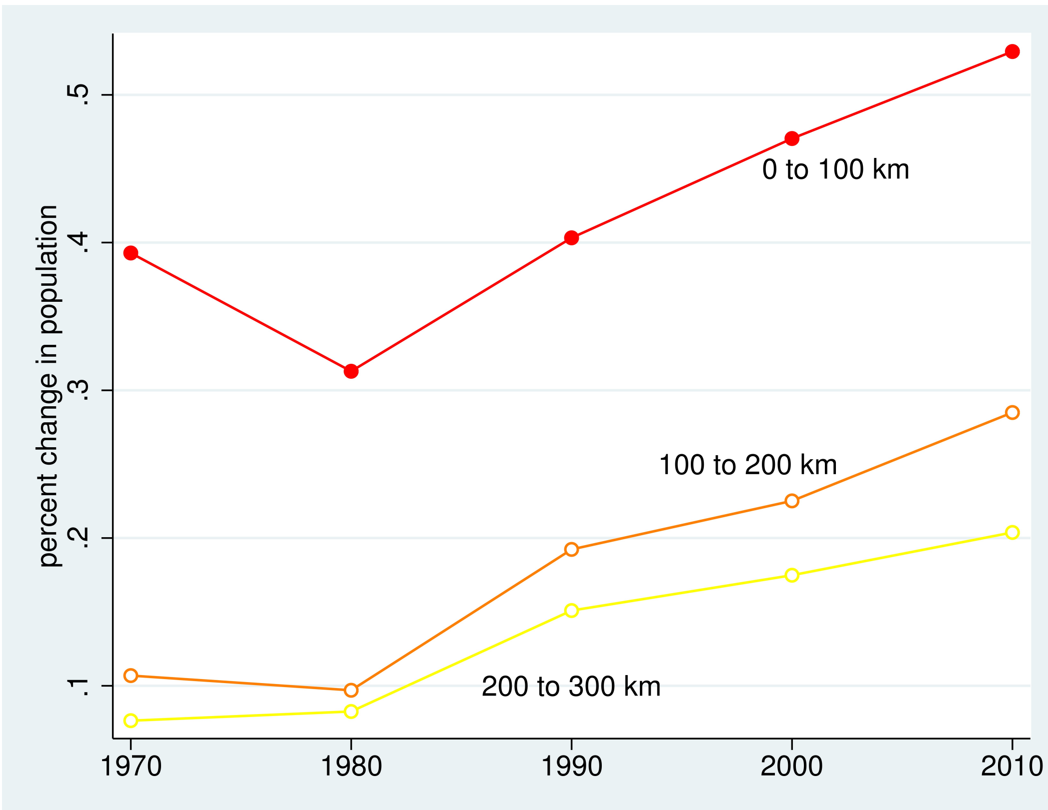
Notes: "Identifying variation" is the residual from a regression of the instrument (county is within $d_1$ to $d_2$ km of a "very deep" port in 1953) on the full set of covariates. Appendix Figures 2a, 2c, and 2e plot the identifying variation versus the residual of a regression of 1910 log population on the full set of covariates from Table 2. Appendix Figures 2b, 2d, and 2f plot the identifying variation versus the residual from a regression of total dollars of 1948 international trade at ports between $d_1$ to $d_2$ km of a county, conditional on the full set of covariates.

Appendix Figure 3: IV Estimates Indistinguishable From Zero at 300 km



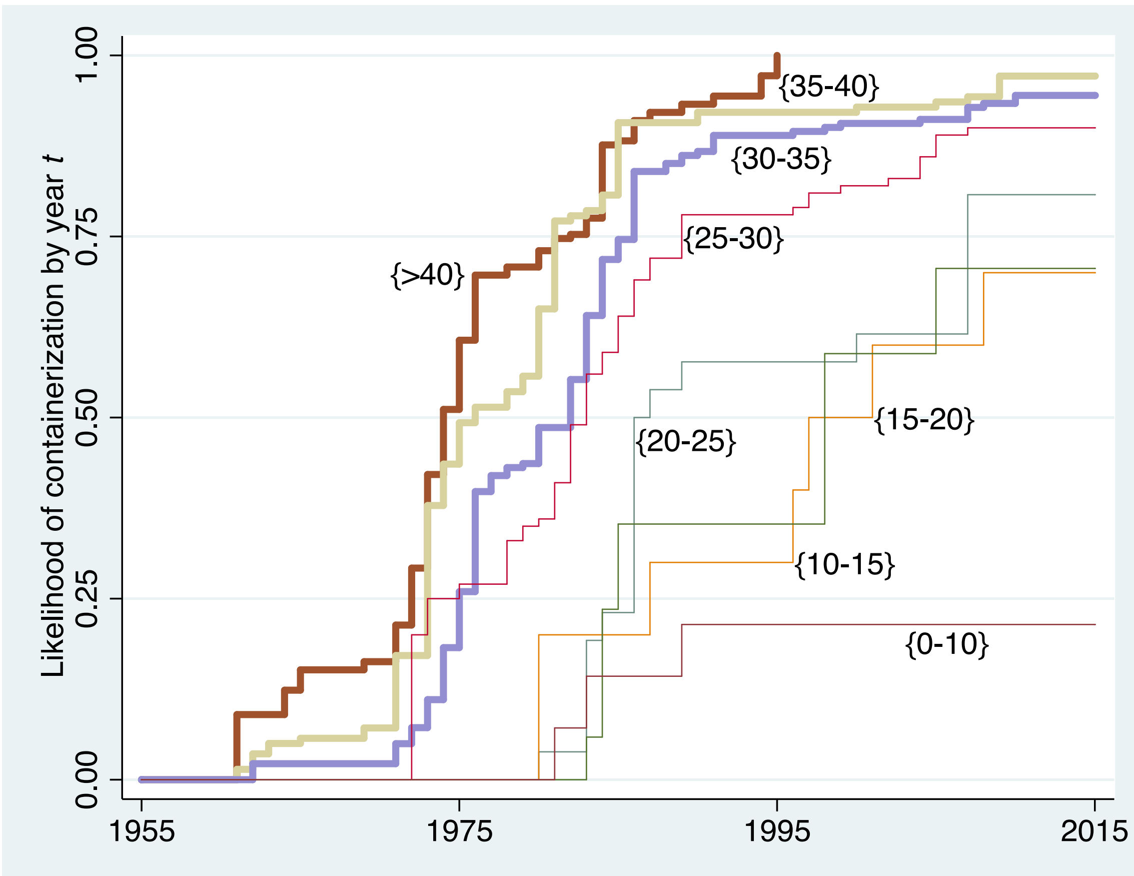

Notes: This picture reports coefficients from the specification in column 8 of Table 2, but paramaterize $\Delta C_{i,t}$ as six indicator variables, one for each distance bin of {0 to 50, 50 to 100, 100 to 150, 150 to 200, 200 to 250, 250 to 300} km. Each dot is an estimated coefficient from this regression and gray bands portray the 90% and 95% confidence intervals.

Appendix Figure 4: Containerization's Impact Increases Over Time




Notes: This picture reports coefficients from the specification in column 8 of Table 2, but where the dependent variable is the change in log population from 1950 to the year reported on the horizontal axis and the endogenous variable is the change in containerization status from 1950 to the year reported on the horizontal axis. Each dot corresponds to an estimated coefficient by distance bin. Full circles are significant at the 5 percent level; hollow circles are insignificant coefficients.

Appendix Figure 5: Depth and Likelihood of Containerization, World Cities




Notes: Lines in the figure report the likelihood that a city will have a containerized port within 300 km in year *t* by the depth of the deepest port within 300 km in 1953. We use thick lines to draw counties near ports that we classify as "very deep," and thin lines for the remainder of cities. The likelihood of being proximate to a container port is greater the closer the city is to a very deep 1953 port.

Appendix Table 1: Complete First Stage Specification

| | 1 if Nearest Container Port is $d_1$ to $d_2$ km of county | | |
|---|---|---|---|
| | 0 to 100 | 100 to 200 | 200 to 300 |
| | (1) | (2) | (3) |
| County is $d_1$ to $d_2$ of a very deep port | | | |
| 0 to 100 km | 0.542*** | 0.068 | -0.012 |
| | (0.067) | (0.066) | (0.046) |
| 100 to 200 km | 0.015 | 0.605*** | -0.013 |
| | (0.034) | (0.049) | (0.042) |
| 200 to 300 km | -0.016 | -0.017 | 0.632*** |
| | (0.027) | (0.04) | (0.052) |
| R-squared | 0.584 | 0.462 | 0.416 |
| Joint F test, instruments | 22.4 | 59 | 56.8 |
| Mean, dependent variable | 0.122 | 0.173 | 0.146 |

Notes: All estimations use 3,023 observations. Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. The F test values in this table are from a test of joint significance of the three reported coefficients. Table 2 reports the Kleinberg-Paap F statistic, as suggested by Sanderson and Windmeijer (2016).

Appendix Table 2: Midwest Counties Have No First Stage and Reduced Form Impacts Are Zero

| | First Stage | | | Reduced Form |
|---|---|---|---|---|
| | 1 if Closest Container Port is $d_1$ to $d_2$ km of county | | | |
| | 0 to 100 | 100 to 200 | 200 to 300 | Change in Log Population, 1950 to 2010 |
| | (1) | (2) | (3) | (4) |
| County is $d_1$ to $d_2$ of a very deep port | | | | |
| 0 to 100 km | 0.065 | -0.332* | 0.248* | -0.04 |
| | (0.146) | (0.188) | (0.146) | (0.158) |
| 100 to 200 km | 0.015 | -0.011 | -0.06 | 0.078 |
| | (0.083) | (0.175) | (0.143) | (0.103) |
| 200 to 300 km | -0.031 | -0.376*** | 0.496*** | 0.05 |
| | (0.073) | (0.137) | (0.106) | (0.081) |
| R-squared | 0.512 | 0.264 | 0.269 | 0.332 |
| Joint F test, instruments | 0.4 | 5.3 | 13.5 | . |
| Mean, dependent variable | 0.16 | 0.329 | 0.35 | 0.397 |

Notes: Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. The sample is restricted to the Midwest Census region, which has no very deep ports. All regressions use 702 observations and cluster standard errors at the 2010 commuting zone.

Appendix Table 3: World City Characteristics by Distance to Nearest Containerized Port

| | Distance to Containerized Port, km | | | | |
|---|---|---|---|---|---|
| | 0 to 100 | 100 to 200 | 200 to 300 | Ever Cont. | Never Cont. |
| | (1) | (2) | (3) | (4) | (5) |
| Log Population | | | | | |
| 1950 | 12.52 | 12.03 | 12.05 | 12.32 | 11.98 |
| | [1.11] | [0.94] | [0.88] | [1.06] | [0.81] |
| 2010 | 13.97 | 13.55 | 13.60 | 13.81 | 13.61 |
| | [1.04] | [0.85] | [0.80] | [0.98] | [0.80] |
| Continent | | | | | |
| Africa | 0.10 | 0.09 | 0.09 | 0.10 | 0.05 |
| Asia | 0.36 | 0.40 | 0.40 | 0.38 | 0.59 |
| Australia | 0.02 | 0.00 | 0.00 | 0.01 | 0.00 |
| Europe | 0.25 | 0.23 | 0.23 | 0.24 | 0.19 |
| North America | 0.18 | 0.19 | 0.20 | 0.19 | 0.11 |
| South America | 0.09 | 0.08 | 0.09 | 0.09 | 0.08 |
| Observations | 373 | 159 | 102 | 634 | 417 |

Note: The unit of observation in this table is a city with at least 50,000 inhabitants in 1950. We report means and standard deviations in brackets. See data appendix for more details on the world sample.

Appendix Table 4: Complete First Stage Estimates for World Sample

| | 1 if Closest Container Port is $d_1$ to $d_2$ km of city | | |
|---|---|---|---|
| | 0 to 100 | 100 to 200 | 200 to 300 |
| | (1) | (2) | (3) |
| City is $d_1$ to $d_2$ of a very deep port | | | |
| 0 to 100 km | 0.573*** | -0.019 | -0.033 |
| | (0.045) | (0.033) | (0.028) |
| 100 to 200 km | 0.020 | 0.579*** | -0.033 |
| | (0.048) | (0.050) | (0.032) |
| 200 to 300 km | 0.006 | 0.099* | 0.511*** |
| | (0.045) | (0.047) | (0.055) |
| R-squared | 0.653 | 0.457 | 0.406 |
| Joint F test, instruments | 75.7 | 58.4 | 37.7 |
| Mean, dependent variable | 0.355 | 0.151 | 0.097 |

Notes: All estimations use 1,051 observations. Stars denote significance levels: * 0.10, ** 0.05, and *** 0.01. The F test values in this table are from a test of joint significance of the three reported coefficients. Table 4 reports the Kleinberg-Paap F statistic, as suggested by Sanderson and Windmeijer (2016).

Appendix Table 5: County Groupings for Consistent Counties

| State | State FIPS | Grouped County FIPS | Initial Counties: County Name | Initial Counties: County FIPS | Notes |
|---|---|---|---|---|---|
| Arizona | 04 | 027 | La Paz County | 012 | Used to be part of Yuma County (04/027) |
| Florida | 12 | 086 | Miami Dade | 025 | Name change, from Dade County to Miami-Dade, yielded a numbering change. |
| Hawaii | 15 | 010 | Kalawao County | 005 | |
| Hawaii | 15 | 010 | Maui County | 009 | |
| Montana | 30 | 067 | Yellowstone County | 113 | Yellowstone County merged is to Park County (30/067) |
| Nevada | 32 | 510 | Ormsby County | 025 | Becomes Carson City (32/510) |
| New Mexico | 35 | 061 | Cibola County | 006 | Used to be part of Valencia County (35/061) |
| South Dakota | 46 | 041 | Armstrong County | 001 | Is merged into Dewey County (46/041) |
| South Dakota | 46 | 071 | Washabaugh County | 131 | Is merged into Jackson County (46/071) |
| Virginia | 51 | 900 | Albermarle County | 003 | |
| Virginia | 51 | 901 | Alleghany County | 005 | |
| Virginia | 51 | 906 | Arlington County | 013 | |
| Virginia | 51 | 902 | Augusta County | 015 | |
| Virginia | 51 | 903 | Bedford County | 019 | |
| Virginia | 51 | 903 | Campbell County | 031 | |
| Virginia | 51 | 904 | Carroll County | 035 | |
| Virginia | 51 | 905 | Chesterfield County | 041 | |
| Virginia | 51 | 915 | Dinwiddie County | 053 | |
| Virginia | 51 | 924 | Elizabeth City | 055 | |

| | | | | | |
|---|---|---|---|---|---|
| Virginia | 51 | 906 | Fairfax County | 059 | |
| Virginia | 51 | 907 | Frederick Couty | 069 | |
| Virginia | 51 | 904 | Grayson County | 077 | |
| Virginia | 51 | 908 | Greensville County | 081 | |
| Virginia | 51 | 909 | Halifax County | 083 | |
| Virginia | 51 | 905 | Henrico County | 087 | |
| Virginia | 51 | 910 | Henry County | 089 | |
| Virginia | 51 | 911 | James City County | 095 | |
| Virginia | 51 | 912 | Montgomery County | 121 | |
| Virginia | 51 | 800 | Nanasemond City | 123 | Is later folded into Suffolk County (51/800) |
| Virginia | 51 | 913 | Norfolk County | 129 | |
| Virginia | 51 | 914 | Pittsylvania County | 143 | |
| Virginia | 51 | 915 | Prince George County | 149 | |
| Virginia | 51 | 913 | Princess Anne | 151 | |
| Virginia | 51 | 916 | Prince William County | 153 | |
| Virginia | 51 | 917 | Roanoake County | 161 | |
| Virginia | 51 | 918 | Rockbridge County | 163 | |
| Virginia | 51 | 919 | Rockingham County | 165 | |
| Virginia | 51 | 920 | Southhampton County | 175 | |
| Virginia | 51 | 921 | Spotsylvania County | 177 | |
| Virginia | 51 | 924 | Warwick County | 189 | |
| Virginia | 51 | 922 | Washington County | 191 | |
| Virginia | 51 | 923 | Wise County | 195 | |
| Virginia | 51 | 924 | York County | 199 | |
| Virginia | 51 | 906 | Alexandria City | 510 | |
| Virginia | 51 | 903 | Bedford City | 515 | |
| Virginia | 51 | 922 | Bristol City | 520 | |
| Virginia | 51 | 918 | Buena Vista City | 530 | |
| Virginia | 51 | 900 | Charlottesville City | 540 | |
| Virginia | 51 | 913 | Chesapeake City | 550 | |
| Virginia | 51 | 901 | Clifton Forge City | 560 | |
| Virginia | 51 | 905 | Colonial Heights City | 570 | |
| Virginia | 51 | 901 | Covington City | 580 | |

| | | | | | |
|---|---|---|---|---|---|
| Virginia | 51 | 914 | Danville City | 590 | |
| Virginia | 51 | 908 | Emporia City | 595 | |
| Virginia | 51 | 906 | Fairfax City | 600 | |
| Virginia | 51 | 906 | Falls Church City | 610 | |
| Virginia | 51 | 920 | Franklin City | 620 | |
| Virginia | 51 | 921 | Fredricksburg City | 630 | |
| Virginia | 51 | 904 | Galax City | 640 | |
| Virginia | 51 | 924 | Hampton City | 650 | |
| Virginia | 51 | 919 | Harrisonburg City | 660 | |
| Virginia | 51 | 915 | Hopewell City | 670 | |
| Virginia | 51 | 918 | Lexington City | 678 | |
| Virginia | 51 | 903 | Lynchburg City | 680 | |
| Virginia | 51 | 916 | Manassas City | 683 | |
| Virginia | 51 | 916 | Manassas Park City | 685 | |
| Virginia | 51 | 910 | Martinsville City | 690 | |
| Virginia | 51 | 800 | Nanasemond County | 695 | Appears for a few years in County Business Patterns data as a county. |
| Virginia | 51 | 924 | Newport News City | 700 | |
| Virginia | 51 | 913 | Norfolk City | 710 | |
| Virginia | 51 | 913 | Portsmouth City | 710 | |
| Virginia | 51 | 923 | Norton City | 720 | |
| Virginia | 51 | 915 | Petersburg City | 730 | |
| Virginia | 51 | 924 | Poquoson City | 735 | |
| Virginia | 51 | 912 | Radford City | 750 | |
| Virginia | 51 | 905 | Richmond City | 760 | |
| Virginia | 51 | 917 | Roanoake City | 770 | |
| Virginia | 51 | 917 | Salem City | 775 | |
| Virginia | 51 | 909 | South Boston City | 780 | |
| Virginia | 51 | 913 | South Norfolk City | 785 | |
| Virginia | 51 | 902 | Staunton City | 790 | |
| Virginia | 51 | 913 | Virginia Beach City | 810 | |
| Virginia | 51 | 902 | Waynesboro City | 820 | |
| Virginia | 51 | 911 | Williamsburg City | 830 | |

| | | | | | |
|---|---|---|---|---|---|
| Virginia | 51 | 907 | Winchester City | 840 | |
| Wyoming | 56 | 039 | Yellowstone Park County | 047 | Is merged into Teton County (56/039) |